

# THIS WEEK

## EDITORIALS

**WORLD VIEW** Ahmed Zewail explains why scientists should not be managed **p.347**

**DRONGOS** Bird study shows crime does pay — for victims **p.349**

**SUPER DUPER** New computer means no more slow bytes from China **p.351**



## Closing the Climategate

*The official inquiry might have exonerated scientists, but attitude changes are needed for science to ensure it holds the public's trust.*

This week marks the first anniversary of the worldwide scandal over the release of e-mails stolen from a computer server at the University of East Anglia (UEA) in Norwich, UK. The server was in the university's Climatic Research Unit (CRU), most of the correspondents involved were climate scientists and the affair will be forever known as Climategate. The scientist at the centre of the storm, Phil Jones, the head of CRU, tells *Nature* on page 362 that he feels the worst is behind him.

It would be naive for Jones and other scientists to assume that the fuss has passed into history. Never mind that almost all of the accusations thrown at the researchers involved have been proven baseless. Never mind that much of the media has retreated from the aggressive stance it adopted during its 'comment first, ask questions later' approach to the content of the e-mails. And never mind that the scientific basis for the global-warming problem remains as solid as it was a year ago. Huge damage has been done to the reputation of climate science, and arguably to science as a whole. That impact deserves to be assessed and the necessary lessons need to be learned.

Take the name Climategate itself. The 'gate' suffix, now routinely applied to the most mundane controversies, is as trite as it is predictable. At the height of the controversy, senior figures called for journalists not to use the word, which they argued lent false seriousness to far-fetched claims of research skulduggery and corruption. That reaction alone helps to explain the sluggish response of the science establishment a year ago to the allegations made against their colleagues and their profession. One lesson that must be taken from Climategate is that scientists do not get to define the terms by which others see them and their place in society. This journal has already warned that climate scientists have to accept that they are in a street fight. They should expect a few low blows. The key is to learn which punches to roll with and which to block and counter.

### TYPICAL EXCHANGES

Take peer review. To many veterans of this bruising process, the talk from Jones in the e-mails of going to town on negative reviews to keep papers from being published was run-of-the-mill stuff. "That's nothing, you should see the rudeness of some of the reviews that go around in microbiology/quantum physics/oncology," was a common reaction. To the outside world, such bravado did little to appease. Likewise, many were shocked by the foolish (if vain) e-mailed boasts of Jones to keep poor papers from inclusion in a report of the Intergovernmental Panel on Climate Change, even if it meant having to "redefine what the peer-review literature is".

The official inquiry into the e-mail affair concluded that such robust exchanges were typical in science. But many non-scientists were still unconvinced. They hold peer review as a revered gold standard of scientific excellence, not to be questioned or used as an opportunity to be rude about academic rivals, even in private. Why? Researchers

may routinely complain about the shortcomings of peer review to other scientists, but they often unite behind it in the face of criticism from outside the scientific sphere. That a study has been through peer review is used too often as a universal defence of its quality. If more scientists were more forthcoming about the flaws in their quality-control

**"Climate scientists have to accept that they are in a street fight. They should expect a few low blows."**

system, then commentators and the wider public may have been more willing to accept that scientists engaged in it do not always act as the public would expect.

With the official inquiry clearing the CRU scientists of fudging data and of abusing the peer-review process, most of the more informed criticism has now settled on the fuzzy notion of the need for greater transparency and openness. Calls for full release of computer code written by climate researchers seem driven more by the fact that it is not routinely made available rather than because it is particularly useful, but it is clear that the CRU scientists did not cooperate fully with all requests for data and other information.

### DUTY TO REPORT

For critics of CRU and their, sometimes legitimate, complaints about data access to be taken seriously, they must be more specific about who should be more open with what, and address their concerns at the correct target. It remains the case that many of the data used by CRU scientists are covered by agreements that prevent their wider distribution. This is not ideal, but it is hardly the fault of the CRU researchers — even if they did seem reluctant to share.

Climate is not the only research area affected by such data restrictions — a paper published earlier this year on the failure of African game parks to conserve large mammals, for example, could not present local data it gathered from reserve operators, who wanted it kept confidential (I. D. Craigie *et al. Biol. Conserv.* **143**, 2221–2228; 2010). There are often good reasons for such sequestering of data, and some studies might not be done without it. But where the full information needed to reproduce a study is not publicly available, scientists have a duty to report that, and say why.

Just as scientists cannot choose the name of future scandals, they cannot choose where allegations will appear. The UEA has taken some justified heavy fire for its handling of the crisis, which was crippled by the enforced absence on medical grounds of Jones, its chief defence witness. Had Jones been strong enough to face the media at the beginning, and say many of the things he says now, the crisis may have blown itself out. The UEA hierarchy misjudged the need to respond and the role that Internet blogs now play in seeding stories for the mainstream media. "I won't worry about it until I hear it on the [BBC Radio] *Today* programme," one university official said when pointed to early online coverage at the time. He got his wish a few days later. By then, the Climategate was already swinging off its hinges. ■

# Scientists wanted

*A clumsy immigration cap could damage UK science by keeping skilled researchers out.*

**A**mong the vacancies for shop assistants and forklift-truck drivers advertised to job-seekers in Hinxton, a village near Cambridge, UK, there are some more specialized positions. A molecular geneticist, for example, is needed to develop scalable technologies for genetic modification of the *Plasmodium falciparum* parasite. A bone biologist is also wanted, with in-depth knowledge of mouse genetics and endocrine systems.

The adverts are for postdoctoral positions at the nearby Wellcome Trust Sanger Institute, a world-class research centre. Traditionally, the institute has not struggled to fill such posts: if no suitable local candidate came forward, it could always recruit from overseas. Science is a global game after all, and talent has no respect for national borders.

The Sanger Institute is among the UK academic and research institutions now threatened by a clumsy cap on immigration, introduced by the Conservative–Liberal Democrat coalition government. Under interim measures in place until the end of March, the number of workers who can enter Britain from outside the European Economic Area has been strictly limited. Positions at UK universities promised to overseas scientists have already been withdrawn. *The Times* newspaper, which has turned a much-needed spotlight on the situation, reports that the cap has already seen more than 230 scientists and academics barred from obtaining the necessary entry visas. Some will be eligible to enter Britain next year. Many will not bother.

The great and the good of British science, many of whom come from overseas or have imported team members, have queued up to warn of the folly of such a policy. In the United States, tighter restrictions on entry for scientists — introduced in response to the terrorist attacks in 2001 — have increased the costs and delays of overseas recruitment, hit international collaborations and been widely viewed as damaging to US science. At a time when nations such as China and Germany are increasing investment in their research bases, Britain is turning away some of the people it needs the most.

There is no evidence that UK Prime Minister David Cameron and his cabinet want to pull up the drawbridge against researchers and erect ‘British science closed’ signs at the airports. But curbs on general immigration were promised by all three major parties prior to this

**“Britain must face the truth: it needs the best scientists more than they need it.”**

year’s election, and the numbers of money-spinning overseas students and those who seek political asylum are harder to restrict than the numbers of skilled workers. The unintended damage to science will be on the agenda later this month, when the cabinet discusses what to do with the cap from April.

An exemption for researchers of a certain calibre (similar to the existing route into Britain for overseas star footballers) is one option, but would exclude promising young scientists who have not yet been able to prove their value. Short of reversing the changes this year that saw, for example, reduced importance given to a PhD in the evaluation of visa applications, the most logical step for the government is to restore the freedom for academic institutions to recruit whoever they wish for more junior positions. If necessary, a trial period could be undertaken, and be scrutinized for abuse. Britain must face an uncomfortable truth: it needs the best scientists more than they need it. ■

# Scope for change

*Tough lessons must be learned if NASA is to avoid repeating a costly accounting error.*

**I**t is hard to keep track of which is expanding faster: the accelerating Universe that the James Webb Space Telescope (JWST) is designed to explore, or the telescope’s cost, which last week inflated from US\$5 billion to \$6.5 billion. Even for NASA, which has a well-documented history of going over budget on major projects, the \$1.5-billion jump is a shocker. Its consequences will surely be felt across the US astronomy community, as well as by the project’s international partners. Even more distressing is the realization that the problems might have been avoided.

According to an independent review (see page 353), NASA administrators did the JWST a significant disservice by concealing its true costs after it was approved. Again and again, they passed overruns to the following year’s budget in a hopeless effort to pay tomorrow for what was needed today. The repeated deferrals mean that the JWST, a tremendously ambitious undertaking by any measure, will now cost US taxpayers far more than it should have done. The report rightly lays most of the responsibility for this at NASA’s door, but Congress deserves a share of the blame. Political wrangling has consistently constrained the space agency’s budget without reducing public and political expectations. As has been seen with planetary exploration and human space flight, this paradox has bred an administrative culture at NASA that discourages realistic budgeting and honest reporting.

The JWST’s saving grace is that it seems to be technically sound. Although still years from completion, it stands to become one of the most productive astronomical observatories in history. The 6.5-metre infrared

telescope is expected to have roughly six times the light-gathering area of its predecessor, the Hubble Space Telescope. Its gold-coated mirror segments are ideal for probing the atmospheres of planets in distant solar systems and reaching back into the early history of the Universe to capture light from the first stars. Given Hubble’s transformational impact on astronomy — and on the wider public’s engagement with science — the case for a next-generation, all-purpose space observatory seems as strong as ever. That makes it all the more urgent to launch the JWST in a timely manner. Once Hubble is retired, the JWST will become the crucial tool with which astronomers can follow up on discoveries made by wide-field survey telescopes on the ground and in space.

But no project, however worthy, is too big to fail, and the JWST has now swerved disturbingly close to fiasco. To keep it to its latest price tag will be a painful process that will damage future projects and further erode the space agency’s credibility. One casualty could be the proposed Wide Field Infrared Space Telescope (WFIRST), a high-priority mission to study the mysterious ‘dark energy’ that seems to pervade the Universe. NASA should think again about this project in light of a proposed European mission that would achieve similar results.

The agency might avoid a wholesale gutting of space astrophysics if it concentrates on small-to-medium-sized missions while it clears the JWST from its books. Such missions are exactly what is proposed in a recent decadal survey by the US astronomy community.

Much harder will be the task of re-engineering NASA to avoid a repeat. The independent review makes specific recommendations, which NASA should pursue in earnest. These include better communication between NASA headquarters and the centres where projects are carried out, as well as more frequent independent reviews, which need to become routine. At its best, NASA allows humanity to look towards the stars. To continue doing so, the agency’s leadership must keep its feet firmly on the ground. ■

**➤ NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunjv](http://go.nature.com/xhunjv)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## ASTRONOMY

### Old galaxies have bars of stars

At least 30% of disk-shaped galaxies, including the Milky Way, have a thick line, or 'bar', of stars, dust and gas across their centre. A new study shows that these bars are more common in older galaxies than younger ones, suggesting that they might cause galaxies to age more quickly.

Karen Masters at the University of Portsmouth, UK, and her colleagues examined data from 13,665 disk galaxies. The galaxies had already been catalogued by the Galaxy Zoo project, which enlists members of the public to comb through telescope data and classify galaxies. The researchers found that as many as half of redder galaxies — which host only older stars — have bars (**pictured**, top). By contrast, 80–90% of bluer galaxies — in which stars are currently being born in large numbers — do not (bottom).

*Mon. Not. R. Astron. Soc.*  
doi:10.1111/j.1365-  
2966.2010.17834.x (2010)

SDSS



## EPIGENETICS

### What makes a queen bee?

A queen honeybee and her female workers have identical DNA sequences but obvious differences in behaviour and reproductive ability. This can be explained, in part, by the attachment of methyl groups to the bees' DNA, which changes gene expression. Now researchers have found significant differences in the methylation patterns of more than 550 genes — most of which are involved in essential cellular activities such as metabolism and RNA synthesis.

Ryszard Maleszka at the Australian National

University in Canberra and his team analysed genomes from the brain tissue of reproductive queens and sterile workers to reveal the genome-wide distribution of methyl groups. They found that methylation sites clustered in areas of genes where splicing — a form of cutting and pasting — occurs in the RNA that is transcribed from the gene. The authors say that methylation may influence the splicing process to generate different gene products.

*PLoS Biol.* 8, e1000506 (2010)

## MATERIALS SCIENCE

### Graphene meets fluorine

The graphene family has a new member: fluorographene, an atom-thick sheet of carbon in which a fluorine atom is attached to every carbon atom. The material is effectively a two-dimensional analogue of Teflon.

Rahul Nair and Andre Geim at the University of Manchester, UK, and their co-workers made the material by exposing a sheet of graphene to fluorine atoms. Unlike graphene, which has high electrical conductivity, fluorographene exhibits high resistance, making it a

candidate insulating material for electronic applications.

With its outstanding thermal and chemical stability, and mechanical properties that exceed those of steel, fluorographene could be used in similar ways to Teflon.

*Small* doi:10.1002/  
smll.201001555 (2010)

## PHARMACOLOGY

### Blocking a gut reaction

Some metabolized cancer drugs are reactivated by bacteria in the gut as they pass through, causing severe diarrhoea. Matthew Redinbo at the University of North Carolina at Chapel Hill and his

colleagues have identified the molecular structure of a key bacterial enzyme responsible, and have found several chemicals that prevent the reactivation.

The small-molecule inhibitors block the active site of the enzyme, called  $\beta$ -glucuronidase, without killing the resident bacteria or harming mammalian cells, which produce a slightly different version of the enzyme. One inhibitor protected mice from diarrhoea and colon damage caused by the drug irinotecan. Similar treatment could allow cancer patients to tolerate higher, more effective doses of chemotherapy, the researchers say.

*Science* 330, 831–835 (2010)

K. WOTHE/BUCKWINKEL/STILL PICTURES



## NANOTECHNOLOGY

## DNA tiles yield bigger arrays

DNA 'origami' describes the practice of using specially designed DNA molecules to guide the assembly of nanostructures into a variety of shapes. Now researchers have used DNA origami 'tiles' to form two-dimensional crystals with edges reaching 2–3 micrometres in length. This should allow larger and more complex structures to be created, say Nadrian Seeman and his colleagues at New York University.

The authors used cross-shaped tiles — made from the DNA strands of the M13 virus — with uneven, or 'sticky' ends. Because the axes of the DNA helical strands were perpendicular, the tiles self-assembled in two dimensions to form arrays. This overcomes problems previously encountered with tiles that assembled mainly in one dimension, the authors say. *Angew. Chem. Int. Edn* doi:10.1002/anie.201005911 (2010)

## CELL BIOLOGY

## Communication key to cancer virus

A virus linked to many human cancers may promote the growth of uninfected neighbouring cells by mediating the transfer of key signalling and gene-regulatory molecules from infected cancer cells. These molecules are packaged in tiny sacs called exosomes, which are taken up by the nearby cells.

Nancy Raab-Traub and her team at the University of North Carolina at Chapel Hill isolated exosomes from cancer cells that had been infected with Epstein–Barr virus. They found that these contained high levels of LMP1, a protein encoded by the virus that enhances cell growth and is found in many cancers. After incubating normal cells with the exosomes, the authors found

LMP1 and activated growth-signalling pathways in the cells. The cells also contained viral microRNAs, which regulate gene expression, suggesting that the virus uses exosomes to manipulate its environment. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1014194107 (2010)

## GENETICS

## Metabolic variation's roots

Metabolism is under the control of a combination of heritable and environmental factors. Teasing out how these factors interact could help to explain why metabolism differs between individuals.

Daniel Kliebenstein at the University of California, Davis, and his colleagues looked for associations between more than 200,000 single-nucleotide variants across the genome and levels of 327 metabolites in 96 strains of *Arabidopsis thaliana*. They found that only 23–30% of the variation in cellular metabolite levels was associated with specific sites in the genome. The team also noted that high genetic variation was not associated with high metabolic variation.

The results suggest that many small gene effects control metabolism, and point to the need to examine metabolism under a range of environmental conditions to fully dissect its genetics. *PLoS Genet.* 6, e1001198 (2010)

## NUCLEAR PHYSICS

## Isotopes map uncharted realm

The production of six new heavy isotopes promises to shed light on the shell model for nuclear structure of the periodic table's heavier elements.

The new nuclei — which fit into the periodic table between rutherfordium (element 104) and the as-yet-unnamed element 114 — were created in a single radioactive decay

## COMMUNITY CHOICE

The most viewed papers in science

## NEUROSCIENCE

## Brain connections have rhythm

**HIGHLY READ**  
on [www.cell.com](http://www.cell.com)  
in October

The number of neuron-to-neuron connections, or synapses, that an animal has is thought to vary from one time of day to another. A team of scientists at Stanford

University in California set out to watch the process in live zebrafish larvae, using time-lapse fluorescence microscopy.

Lior Appelbaum, currently at Bar-Ilan University in Ramat Gan, Israel, and his colleagues followed the creation and disappearance of synapses over a 24-hour period. They focused on a particular class of neuron in two brain areas known to be involved in regulating sleeping and waking: the pineal gland and the hindbrain.

The authors showed that the number of synapses fluctuated rhythmically between day and night. They also found that a protein, Nptx2, for which levels in the brain also vary rhythmically during the 24-hour period, is involved in regulating the rhythmicity of synapse number.

*Neuron* 68, 87–98 (2010)

chain by Paul Ellison at the Lawrence Berkeley National Laboratory in California and his co-workers. They made the isotopes by hitting targets of plutonium-242 with an intense beam of calcium-48 nuclei, setting off a chain of decays from a nucleus of element 114.

The isotopes' lifetimes ranged from eight-thousandths of a second to just over three minutes. Creating such short-lived isotopes was necessary to generate several examples of the heaviest elements before the nuclei fissioned into two similar-sized parts.

*Phys. Rev. Lett.* 105, 182701 (2010)

## EVOLUTIONARY BIOLOGY

## Food thieves offer a helping hand

An African bird that robs other species of their food seems to help as well as hinder, allowing one of its victims to catch more prey.

Pied babblers (*Turdoides bicolor*; pictured left) are often fooled by fork-tailed

drongos (*Dicrurus adsimilis*; pictured right), which mix fake alarm calls with genuine warnings to distract the babblers, then make off with their food. Andrew Radford at the University of Bristol, UK, and his team studied the birds in the wild, and played back drongo recordings to babblers. They found that babblers captured more prey when reassured by the regular 'twank' noises of real or recorded drongos than when there were no drongos standing guard.

This 'sentinel' calling probably arose as a means of manipulating babblers, but the benefits it brings to both species may mean that the birds' relationship is transitioning from parasitic to mutualistic, the authors suggest.

*Evolution* doi:10.1111/j.1558-5646.2010.01180.x (2010)

➔ **NATURE.COM**

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



A. RADFORD





# SEVEN DAYS

The news in brief

## FUNDING

### Clean-energy cash

A large fund for clean-energy projects in Europe, estimated at €4.5 billion (US\$6.2 billion), launched its first call for proposals on 9 November. The fund, agreed by European member states in February, aims to support at least eight demonstration projects to capture carbon dioxide and store it underground. It will also cover at least 34 projects involving innovative technology for renewable sources such as solar power, bioenergy and wind, tidal and geothermal energy. Money will be raised by selling 300 million carbon credits from the European Union's emissions trading scheme for greenhouse gases; the first projects will be chosen in 2012.

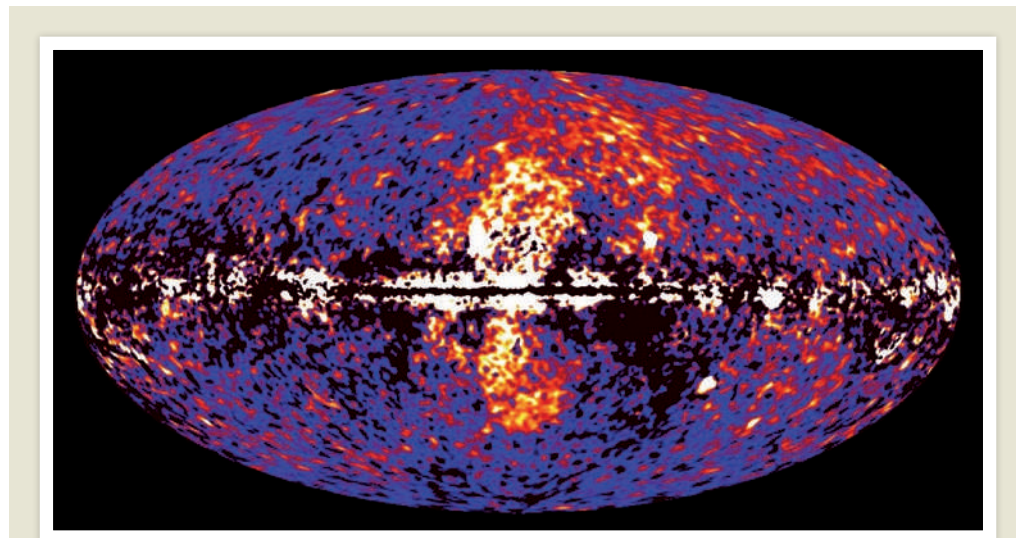
### Telescope woes

The James Webb Space Telescope will cost at least US\$6.5 billion — well over a previous \$5-billion estimate — according to an independent review released on 10 November. See page 353 for more.

## POLICY

### Science statistics

The slow decline of traditional science superpowers was analysed in two statistical reports published last week. The United Nations Educational, Scientific and Cultural Organization (UNESCO) in Paris put out its quinquennial science report ([go.nature.com/gqlokww](http://go.nature.com/gqlokww)), noting that China, India and Korea's share of world research investment and researchers was rising relative to those of the European Union, Japan and the United States. Information company Thomson Reuters of New York City, meanwhile, released a



NASA/DOE/FERMI LAT/D. FINKBEINER ET AL

## Milky Way's double bubble

Using data from NASA's Fermi Gamma-ray Space Telescope, a team of astronomers declared last week that they had discovered two gargantuan 'bubbles' of  $\gamma$ -ray-emitting particles extending north and south of our Galaxy's centre (M. Su *et al. Astrophys. J.* **724**, 1044–1082; 2010). Researchers think the structures, which

measure 15,625 parsecs (50,000 light years) from end to end, formed from a single relatively rapid release of energy equivalent to that from 100,000 supernovae. The source might have been the birth and death of short-lived, massive stars, or a jet of energetic particles from the black hole at the Galactic Centre.

report on the United States as part of a regular series profiling nations ([go.nature.com/skjoe8](http://go.nature.com/skjoe8)). It noted the country's continued strength but waning influence in terms of scientific spending and publications.

### Rewarding impact

UK universities must prepare for their research to be judged on its social and economic benefits, not just its quality, to gain funding. A year-long pilot study testing whether peer-review panels could judge the 'impact' of research was released on 11 November and concluded that the system was workable and robust. See page 357 for more.

### Climate media

Climate science received only token coverage as journalists documented the 2009 United Nations climate summit

in Copenhagen, according to an analysis released on 15 November by the Reuters Institute for the Study of Journalism at the University of Oxford, UK ([go.nature.com/htydhk](http://go.nature.com/htydhk)). Researchers analysed more than 400 articles in 12 countries, and found that nearly 80% of them mentioned climate science in less than 10% of their space. Just 9% of stories mentioned climate science in more than 50% of their space.

## RESEARCH

### First asteroid dust

The Hayabusa space explorer has picked up dust from the Itokawa asteroid, from which it returned in June after a seven-year mission. Researchers at the Japan Aerospace Exploration Agency (JAXA) announced

on 16 November that analysis of the mineral compositions of some 1,500 micrometre-sized grains recovered from Hayabusa's capsule showed that almost all the dust was extraterrestrial and came from Itokawa. This is the first material ever returned to Earth from an asteroid.

### Dengue control

The release of male mosquitoes genetically engineered to be sterile can control dengue fever by suppressing the population of the insects that carry the disease, scientists at Oxitec, a UK-based company part-owned by the University of Oxford, told reporters on 11 November. They were reporting the results of a field trial of transgenic *Aedes aegypti* mosquitoes in a town on Grand Cayman, an

island in the Caribbean Sea. Malaysia will begin field trials of the mosquitoes in the next few months. See [go.nature.com/6rxdjp](http://go.nature.com/6rxdjp) for more.

## Rice research

The world's leading rice-research institutions are joining forces to improve rice yields and breed better varieties. A 5-year, US\$600-million initiative, the Global Rice Science Partnership, was officially launched on 10 November at the third International Rice Congress in Hanoi. It is led by the International Rice Research Institute, based in Los Baños, the Philippines, and part of a consortium of leading agricultural research centres called the Consultative Group on International Agricultural Research. Most of the funding is not new money: current budgets from the centres will be reoriented towards the initiative's research goals. See [go.nature.com/9xjoro](http://go.nature.com/9xjoro) for more.

## Fastest computer

China now possesses the world's speediest supercomputer. As expected, its Tianhe-1A computer (pictured), housed in the National Supercomputer Center in Tianjin, has eclipsed the US Department of Energy's Jaguar system at the Oak Ridge National Laboratory in Tennessee. In the latest update



to the list of the world's top 500 supercomputers ([www.top500.org](http://www.top500.org)), released on 11 November, Tianhe-1A was shown to have achieved 2.57 petaflops ( $2.57 \times 10^{15}$  floating point operations per second), with Jaguar managing 1.75 petaflops. The United States still boasts five of the world's top ten fastest computers.

## Reactome retraction

A hotly debated research paper that described a device called a 'reactome array' able to take rapid snapshots of enzyme activity in a cell (A. Belouqui *et al. Science* **326**, 252–257; 2009) has been retracted by its authors. The retraction was recommended in July by an institutional ethics committee investigation. See [go.nature.com/32lxii](http://go.nature.com/32lxii) for more.

## Ape deaths solved

Japan's premier primate research centre says it has identified the cause of the mysterious series of deaths of its Japanese macaques (*Macaca fuscata*) that had puzzled researchers and

worried citizens earlier this year (see *Nature* **466**, 302–303; 2010). The Primate Research Institute of Kyoto University reported on its website on 11 November that the culprit was simian retrovirus-4 (SRV-4). The problem emerged when the institute housed southeast Asian crab-eating macaques (*Macaca fascicularis*), which are natural carriers of the virus, with Japanese macaques. The report said the virus had never been passed to humans.

### EVENTS

## Cholera in Haiti

The escalating cholera epidemic in Haiti had claimed more than 900 lives and caused close to 15,000 infections by the start of this week, according to the Haitian Ministry of Public Health and Population. The cholera strain is most closely related to one from south Asia, the US Centers for Disease Control and Prevention in Atlanta, Georgia, has said, although it has not pinpointed the source.

### BUSINESS

## Genome market

Complete Genomics, one of the handful of young US companies offering fast, cheap genome sequencing, completed its initial public offering (IPO) on 11 November — raising

## COMING UP

### 20 NOVEMBER

US President Barack Obama's bioethics advisers reach their six-month deadline for completing recommendations on issues raised by synthetic biology. The presidential commission holds its fourth and final public meeting on synthetic biology at Emory University in Atlanta, Georgia, on 16–17 November. [go.nature.com/prqng2](http://go.nature.com/prqng2)

### 21–24 NOVEMBER

Officials from 13 countries with wild tiger populations meet at a global summit on conservation of the species in St Petersburg, Russia. [www.globaltigerinitiative.org](http://www.globaltigerinitiative.org)

US\$54 million at \$9 per share, short of the \$86-million target it set when first filing for an IPO in July. The company, based in Mountain View, California, says it has sequenced more than 400 complete human genomes this year alone. On its first day of public trading, its share price fell 11%.

## Fraud investigation

The European Anti-Fraud Office in Brussels confirmed to *Nature* last week that it is investigating the alleged misuse of European research money by a group of Greek academics — although the agency would not comment on any details of the allegations. An article in the Greek weekly newspaper *Proto Thema* on 7 November reported that up to 20 professors have been accused of embezzling up to €200 million (US\$273 million). See [go.nature.com/qrvykn](http://go.nature.com/qrvykn) for more.

➔ [NATURE.COM](http://NATURE.COM)

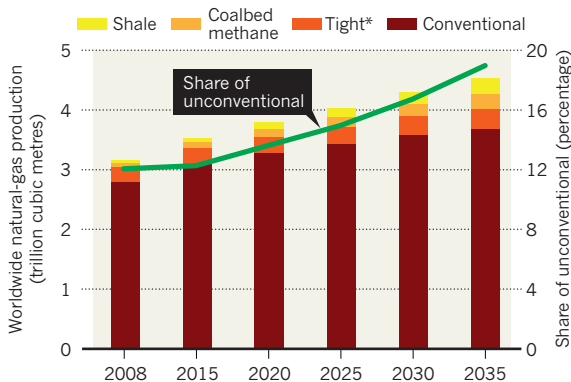
For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

## BUSINESS WATCH

Firms that tap unconventional natural-gas sources — such as in underground shale — are in demand. Atlas Energy of Philadelphia, Pennsylvania, is the latest to be snapped up; on 9 November, oil group Chevron of San Ramon, California, said it would buy the firm in a US\$4.3-billion deal. The moves are based on a belief that such sources will make up an increasing share of global gas production, as projected in the 2010 World Energy Outlook, released on 9 November (see chart).

## BETS PLACED ON NATURAL GAS

Unconventional gas sources are projected to make up an increasing share of rising natural-gas demand.



\*Reservoirs with low permeability to gas flow

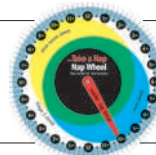


# NEWS IN FOCUS

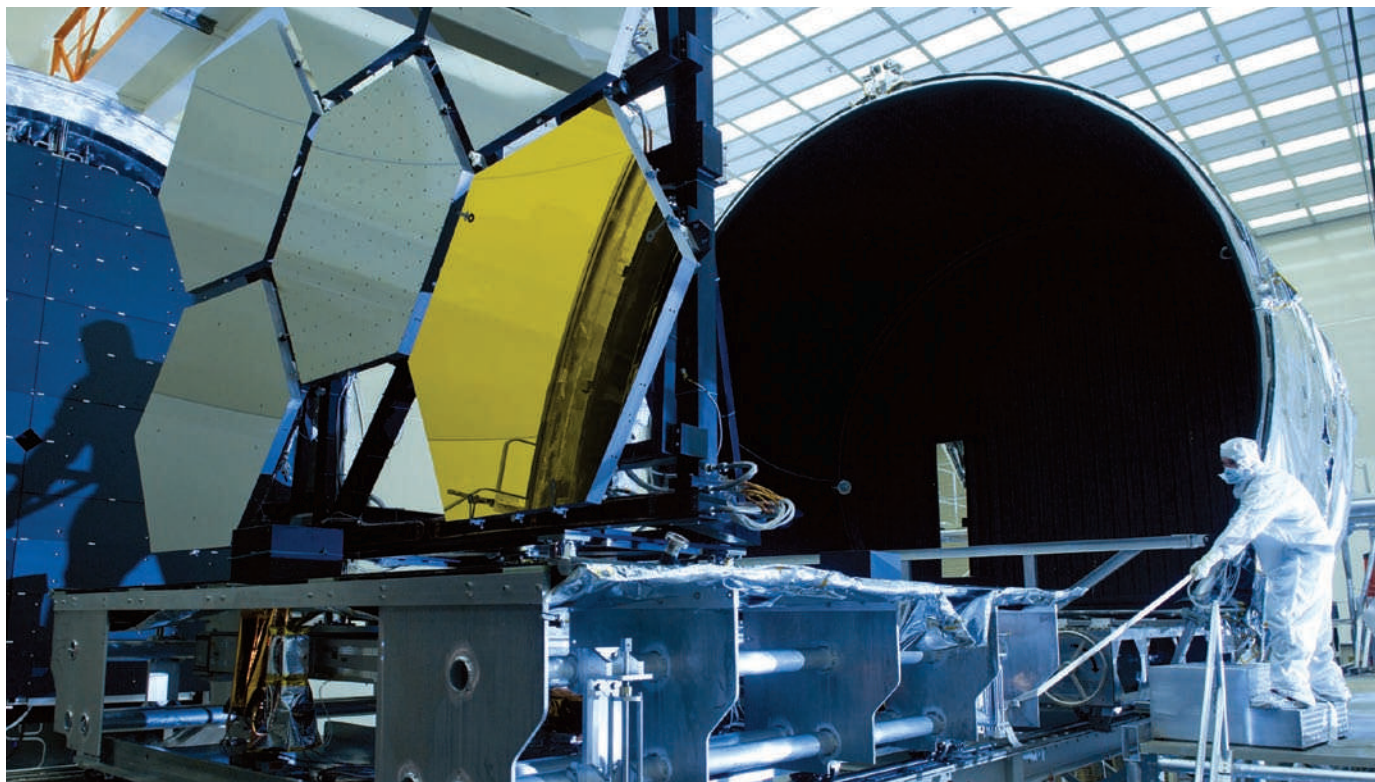
**PHYSICS** Antihydrogen atoms are rounded up for questioning **p.355**

**GERMANY** Network of national centres could boost medical research **p.358**

**CLIMATE** Scientist at heart of e-mail storm has few regrets **p.362**



**MEDIA** The joys and burdens of celebrity in sleep research **p.365**



NASA/MSFC/E. GIVEN

The first of 18 gold-coated, hexagonal mirror segments destined for the James Webb Space Telescope undergoes cryogenic testing.

## ASTRONOMY

# Scope sails into budget void

*An independent review finds NASA's flagship James Webb observatory is perilously overspent.*

BY ADAM MANN

At least no one says it won't work. But that may be the only consolation for NASA administrators as they absorb the implications of a scathing report detailing the budget woes of the James Webb Space Telescope (JWST).

Intended to replace the Hubble Space Telescope, the JWST was estimated to cost US\$1 billion in 2001, but by the time the project received its official go-ahead in 2008, its growing complexity had pushed that figure up to \$5 billion. Now an independent review, released on 10 November, has found that the telescope's true price tag is at least \$6.5 billion,

and that its target launch date has slipped by more than a year to September 2015.

The staggering overrun means that the JWST will have to commandeer vast and unexpected resources, amounting to hundreds of millions of additional dollars per year, to stay on track, officials admitted at a news briefing held to coincide with the report's release. This will undoubtedly have an impact on other projects, first within the affected astrophysics community, then the NASA science programme, and then across the entire agency, said Chris Scolese, associate administrator at

NASA headquarters in Washington DC.

The review, commissioned in June by Senator Barbara Mikulski (Democrat, Maryland) and led by John Casani, an engineer at the Jet Propulsion Laboratory in Pasadena, California, also recommended an administrative reorganization of the project, to which NASA has already acceded. Richard Howard, NASA's deputy chief technologist, will become manager of a new JWST project office based at NASA headquarters, rather than at the Goddard Spaceflight Center in Greenbelt, Maryland, which has been the project's home until now. Howard's first step will be to develop a realistic budget for the JWST, expected by February 2011.

"We do not want to have any more ►

**NATURE.COM**  
For more on the  
JWST, visit:  
[go.nature.com/vy7sy8](http://go.nature.com/vy7sy8)



► surprises in this programme,” Howard told journalists at the briefing.

The report found that the JWST budget presented to NASA in 2008 “was not based upon a current, bottom-up estimate of projected costs” and therefore understated the project’s real requirements. The agency also had inadequate monetary reserves each year to cover expenses that arose when costs turned out to be higher than estimated. NASA dealt with the problem by continually postponing necessary work so that it would fall under a subsequent year’s budget, with the delay causing the work to double or triple in price.

The report admonishes NASA managers, saying that they were aware of the practice of deferring work into future years and “tacitly condoned it”. For some, the findings are evidence that the agency has not got to grips with its tendency to allow programmes to overrun their budgets, which inevitably means that money is siphoned from other projects. “We seem to be better at observing lessons, rather than learning from them,” says Matt Mountain, director of the Space Telescope Science Institute in Baltimore, Maryland.

The JWST will need about \$250 million per year of extra funds in 2011 and 2012, and such estimates represent the minimum, says Casani. Given that the newly elected Congress wants to reduce spending, the extra funds might be unavailable and the launch date would have to be postponed, driving up costs still further, he adds.

The grim outlook limits NASA’s ability to carry out recommendations from the Astro2010 decadal survey, a community-wide effort to assign priorities to major projects, says Alan Stern, a planetary scientist at the Southwest Research Institute in San Antonio, Texas.

“It seems that there was no need for NASA to participate in the decadal, as there are unlikely to be any funds available before 2020 to start anything big and new,” says Alan Boss, chair of the NASA advisory council astrophysics subcommittee and an astrophysicist at the Carnegie Institution for Science in Washington DC. Particularly vulnerable, says Stern, is the Wide-Field Infrared Survey Telescope (WFIRST), the decadal survey’s top large-scale, space-based project. The mission, intended to study the ‘dark energy’ driving the acceleration of the Universe’s expansion, is estimated to cost \$1.6 billion.

Despite its dysfunctional financing, the JWST is technically sound and should still proceed, the report finds, a view widely shared in the community. Mountain points out that it required nearly \$6 billion in today’s dollars to get the Hubble telescope working as it was intended, and few would argue that the money wasn’t well spent.

“It’s good to do one hard project a decade; it reminds us what revolutionary things look like,” he says. ■ [SEE EDITORIAL P.346](#)

## MEDICINE

# Good news for ‘good’ cholesterol

*Positive results inject life into strategy to treat heart disease.*

BY ALLA KATSNELSON

A strategy for lowering heart-disease risk that once seemed to be a dead end is showing fresh promise. Decades of animal studies and epidemiological data had suggested that raising blood levels of high-density lipoprotein — HDL, or ‘good’ cholesterol — might have a stronger protective effect against heart disease than statins, drugs that lower levels of low-density lipoprotein (‘bad’ cholesterol or LDL). But in 2006, a US\$1-billion trial of torcetrapib, an HDL-raising drug, found it seemed to increase patients’ risk of death, casting a pall of doubt over the entire field. This week, the first study since to focus on the class of drugs that boosts HDL levels may offer good news for the approach.

The study, published in *The New England Journal of Medicine*<sup>1</sup>, was a 1,623-patient trial investigating the safety of anacetrapib, a drug functionally similar to torcetrapib, developed by pharmaceuticals giant Merck, based in Whitehouse Station, New Jersey. The drug inhibits a protein called CETP, which raises HDL. The trial found with 94% confidence that anacetrapib does not harm patients — in contrast to the 15,000-patient trial of torcetrapib, also a CETP inhibitor. When Pfizer halted that trial early<sup>2</sup>, many companies stopped working on CETP blockers. Researchers were left wondering whether torcetrapib’s failure was down to unexpectedly high toxicity in that compound, whether the inhibition of CETP itself is harmful, or whether the idea that raising HDL levels lowers risk is flawed.

The anacetrapib trial also tracked the drug’s effects on LDL and HDL levels, which, according to Christopher Cannon, a cardiovascular researcher at Brigham and Women’s Hospital in Boston, Massachusetts, and the study’s principal investigator, are “jawdropping”. After 24 weeks on the drug, patients

experienced a 138% increase in HDL levels. In contrast, exercising and changing diet might only raise HDL by 10%, says Cannon. The participants, all of whom were also on statins, experienced a further 40% reduction in LDL levels.

Although the study wasn’t large enough to look at the effect of anacetrapib on heart disease, the researchers noted some positive trends: 3.3% of patients taking the drug experienced heart attacks, stroke or other kinds of cardiovascular events, compared with 5.3% of patients in the placebo group.

The drug’s apparent safety is encouraging, says Prediman Shah, director of cardiology and atherosclerosis research at Cedars-Sinai Medical Center in Los Angeles, California, “but there are some interesting red flags”. One, he says, has to do with c-reactive protein (CRP), a marker of inflammation in blood that tends to drop as patients regulate their cholesterol with statins or lifestyle changes. Despite the huge changes in LDL and HDL levels, CRP levels actually increased slightly.

Shah also says he is surprised that such enormous shifts in HDL levels yielded such small clinical benefits. “In all fairness, the study wasn’t powered to test that,” says Shah, but if epidemiological predictions on HDL’s benefits are correct, the drug should virtually “confer immortality”.

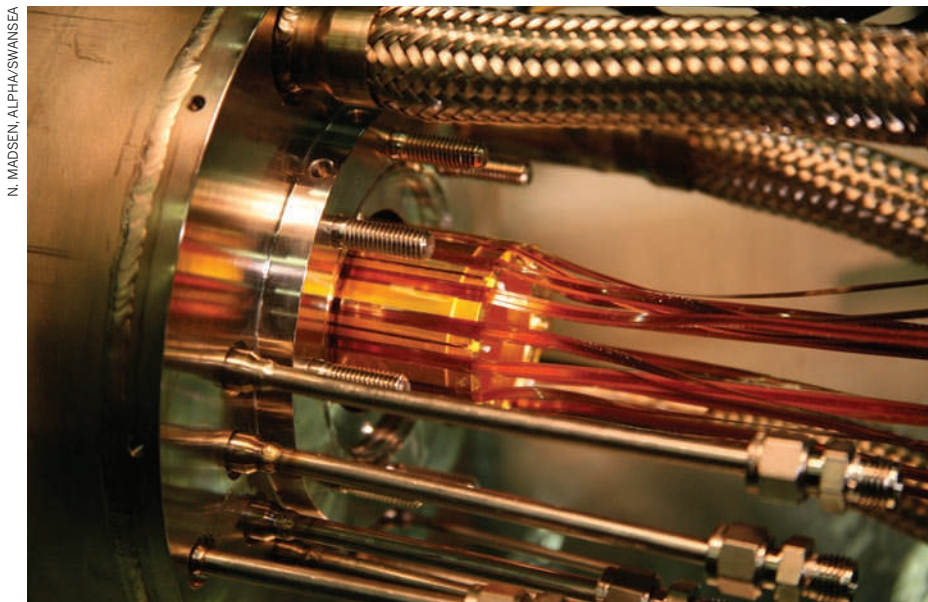
Whether raising HDL really works won’t become clear until data from larger studies begin to emerge, Shah says. An international, 30,000-patient trial testing anacetrapib’s efficacy will begin next year, but results won’t come in until at least 2014 (see Table). Meanwhile, results on another CETP inhibitor called dalcetrapib, developed by Roche, are expected in 2013. ■

1. Cannon, C. P. *et al.* *N. Engl. J. Med.* doi:10.1056/NEJMoa1009744 (2010).

2. Pearson, H. *Nature* **444**, 794–795 (2006).

## TESTING THE EFFICACY OF RAISING LEVELS OF HIGH-DENSITY LIPOPROTEIN

Drug/class	Company	Clinical phase	Results expected
Anacetrapib (CETP inhibitor)	Merck	Enrolling phase III early 2011	2014–15
Dalcetrapib (CETP inhibitor)	Roche	Phase III	2013
Niacin + MK-0524A (to control niacin’s side effects)	Merck	Phase III	2012
RVX-208 (stimulates production of apoA-1, a key HDL protein)	Resverlogix	Phase IIb	2014



The electrodes (gold) of the trap used to combine positrons and antiprotons to form antihydrogen.

# PARTICLE PHYSICS

# Antimatter held for questioning

*Magnetically trapped atoms could test fundamental physics.*

BY EUGENIE SAMUEL REICH

For physicists, a bit of antimatter is a precious gift indeed. By comparing matter to its counterpart, they can test fundamental symmetries that lie at the heart of the standard model of particle physics, and look for hints of new physics beyond. Yet few gifts are as tricky to wrap. Bring a particle of antimatter into contact with its matter counterpart and the two annihilate in a flash of energy.

Now a research collaboration at CERN, Europe's particle-physics lab near Geneva, Switzerland, has managed, 38 times, to confine single antihydrogen atoms in a magnetic trap for more than 170 milliseconds. The group reported the result in *Nature* online on 17 November<sup>1</sup>. "We're ecstatic. This is five years of hard work," says Jeffrey Hangst, spokesman for the ALPHA collaboration at CERN.

An antihydrogen atom is made from a

negatively charged antiproton and a positively charged positron, the antimatter counterpart of the electron. The objective — both for ALPHA and for a competing CERN experiment called ATRAP — is to compare the energy levels in antihydrogen with those of hydrogen, to confirm that antimatter particles experience the same electromagnetic forces as matter particles, a key premise of the standard model. "The goal is to study antihydrogen and you can't do it without trapping it," says Cliff Surko, an antimatter researcher at the University of California, San Diego. "This is really a big deal."

The ALPHA claim is the first major advance since the creation of thousands of antihydrogen atoms in 2002 by a forerunner experiment called ATHENA<sup>2</sup> and by ATRAP<sup>3</sup> (see 'A brief history of antimatter'). Both experiments combined decelerated antiprotons with positrons at CERN to produce antihydrogen atoms. But, within several milliseconds, the atoms

annihilated with the ordinary matter in the walls of their containers.

To prevent that from happening, the ALPHA team formed antihydrogen atoms in a magnetic trap. Although not electrically charged like antiprotons and positrons, antihydrogen — like hydrogen — has a more subtle magnetic character that arises from the spins of its constituent particles. The ALPHA researchers used an octupole magnet, produced by the current flowing in eight wires, to create a magnetic field that was strongest near the walls of the trap, falling to a minimum at the centre, causing the atoms to collect there. To trap just 38 atoms, the group had to run the experiment 335 times. "This was ten thousand times more difficult" than creating untrapped antihydrogen atoms, says Hangst — ATHENA made an estimated 50,000 of them in one go in 2002. To do spectroscopic measurements, Surko estimates that up to 100 antihydrogen atoms may need to be trapped at once.

ATRAP still hopes to reach that goal first. In a paper due out in *Physical Review Letters*, the collaboration reports that it has efficiently separated antiprotons from the cold electrons that are used to cool them down, a step towards creating slower-moving antihydrogen atoms that might stay trapped for longer. "Rather than trying to demonstrate that we can confine 38 antihydrogen atoms for a small fraction of a second, we are working on new methods to produce and trap much larger numbers of colder atoms," says Gerald Gabrielse, ATRAP's spokesman. "We shall see which approach is more fruitful."

Two other collaborations aim to study antihydrogen. In 2003, the international ASACUSA experiment at CERN proposed a scheme to create a beam of antihydrogen atoms<sup>4</sup>. Yasunori Yamazaki, an atomic physicist at the Advanced Science Institute in Saitama, part of Japan's RIKEN network of research labs, now says the group has produced such a beam and may be able to use it to study the energy levels in antihydrogen without needing to trap the atoms. Another CERN experiment called AEGIS is starting to compare the effect of gravity on antihydrogen with that on ordinary hydrogen. Antimatter is almost certain to fall at the same rate as normal matter, but if it doesn't the results could help scientists to distinguish between alternative approaches to unifying quantum theory with general relativity. ■

1. Andresen, G. B. *et al.* *Nature* advance online publication doi:10.1038/nature09610 (2010).
2. Amoretti, M. *et al.* *Nature* **419**, 456–459 (2002).
3. Gabrielse, G. *et al.* *Phys. Rev. Lett.* **89**, 213401 (2002).
4. Mori, A. & Yamazaki, Y. *Europhys. Lett.* **63**, 207–213 (2003).

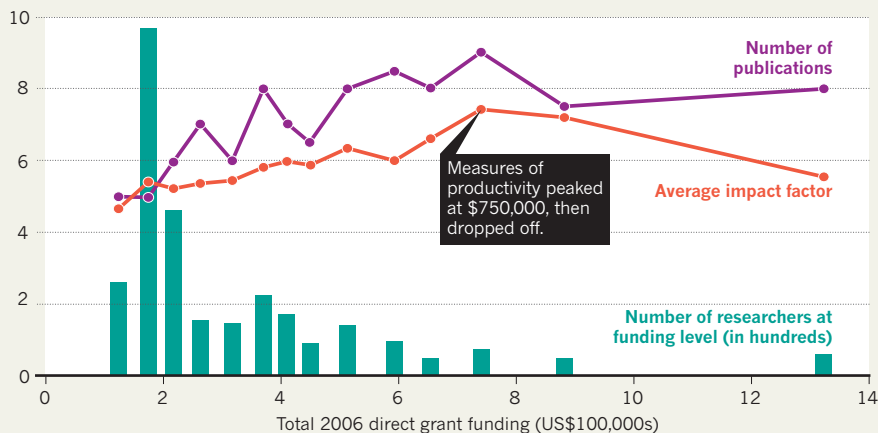
## A BRIEF HISTORY OF ANTIMATTER

1931 ▶	1932 ▶	1955 ▶	1956 ▶	1965 ▶	1995 ▶	2002 ▶	2010
Paul Dirac predicts the existence of antimatter.	Positron discovered at the California Institute of Technology.	Antiproton discovered at Lawrence Berkeley National Laboratory (LBNL) in California.	Antineutron discovered by scientists at the LBNL.	Antideuteron (an antiproton and an antineutron) created at CERN near Geneva, Switzerland, and at Brookhaven National Laboratory in New York.	Observations of antihydrogen at CERN.	Creation of thousands of antihydrogen atoms at CERN.	38 atoms of antihydrogen trapped at CERN.



## MERIT IN THE MIDDLE?

Plotting the median number of grant-linked publications (2007 to mid-2010) and median average journal impact factors against total US National Institutes of Health funding to investigators in 2006 shows the highest performance at medium funding levels.



## METRICS

# Study says middle sized labs do best

*A comparison of funding level and output has captured attention at the US National Institutes of Health.*

BY MEREDITH WADMAN

The director of one of the biggest institutes at the US National Institutes of Health (NIH) posted a blog entry that got tongues wagging this autumn. Jeremy Berg, who heads the National Institute of General Medical Sciences (NIGMS) in Bethesda, Maryland, had analysed the scientific productivity of nearly 3,000 researchers who were funded by grants from his institute in 2006. With the help of NIH data-mining experts, who have developed powerful tools for such studies, Berg was able to show, in hard numbers, what scientists could once only speculate about: the relationship between grant size and scientific productivity.

"Everything had come together so that it seemed possible to ask the questions I asked without it being a two-year project," says Berg.

His analysis plots the median number of publications between 2007 and mid-2010, and the median average impact factor of those publications, against total direct NIH funding in 2006. It covers 2,938 investigators, who were divided into 14 groups on the basis of their funding level.

The resulting plot (see chart) shows that both measures peaked at around US\$750,000

in annual funding; at higher funding levels, the median publication number and average impact factor were both discernibly lower.

Berg says conventional wisdom has long held that, once a lab reaches a certain size, it becomes harder to manage and the average number of publications per dollar falls. But until now, he says, "no one actually had the data to put that in more quantitative terms". He hastens to add that the variation within funding levels is large. "Some people with \$800,000 or \$900,000 are publishing 40 or 50 papers over this time. It's important not to forget that the average behaviour is not the behaviour of everybody."

Berg's analysis comes at a time of increasing austerity for the US government, driven by a struggling economy and ballooning deficits. The push to trim costs is likely to gain strength come January, when spending-conscious Republicans will take control of the US House of Representatives, where funding bills are born. And political cost-cutters may increasingly turn to analyses such as Berg's to inform their decisions.

"Science is not an obvious first choice for the public. It could be regarded as a luxury during a time of recession. So there is a call for greater accountability and greater documentation of the impact and expenditure of public funds," says John Marburger, vice-president for research at

the State University of New York, Stony Brook. As director of the White House Office of Science and Technology Policy under former president George W. Bush, Marburger pushed for more rational systems of developing and evaluating science policy. "Congress and the administration want to see something more than just our anecdotal success stories," adds John McGowan, deputy director for science management at the NIH's National Institute of Allergy and Infectious Diseases in Bethesda.

Analyses similar to Berg's are under way, but on a larger scale. The STAR METRICS (Science and Technology in America's Reinvestment — Measuring the Effects of Research on Innovation, Competitiveness and Science) project was launched in May and, led by the NIH and the US National Science Foundation, aims to develop measurements of the economic and social impacts of US research spending by linking data on federal grant recipients to outcomes such as publications, patents, citations and employment (see *Nature* 464, 488–489; 2010). Meanwhile, McGowan and his team have developed e-SPA (electronic Scientific Portfolio Assistant), a computer tool for gauging productivity by linking NIH-funded investigators to measures including impact factor, citation number and patents applied for and published. e-SPA is now in use by about 1,000 NIH staff as they plan and evaluate their research portfolios and make close-call funding decisions on individual grants. And in 2006, the National Institute of Environmental Health Sciences in Research Triangle Park, North Carolina, launched SPIRES (Scientific Publication Information Retrieval and Evaluation System), an NIH-wide system that matches 275,000 NIH grants with publications going back to 1980.

Some are sceptical of such efforts. "There's no reason to think that just because there is productivity in an area of science it would be a predictor of social value," says Daniel Sarewitz, Washington DC-based co-director of the Consortium for Science, Policy and Outcomes at Arizona State University. "You can be productive on a question that's of great interest to scientists, but of no particular value in terms of application."

Nonetheless, such analyses focus the attention of scientists competing for increasingly scarce dollars. For Dorothy Erie, an NIGMS-funded biochemist at the University of North Carolina in Chapel Hill, Berg's analysis tells an important story. "There's a very clear difference in productivity between those who are above



**"It's important not to forget that the average behaviour is not the behaviour of everybody."**

Jeremy Berg



\$225,000 and those who are below it," she says. "If you can only afford to hire two people, it's hard to be productive."

Berg stresses that the analysis is a conversation-starter, not a judgement to be applied mechanically. "If you just say, 'Based on your funding level, you should be publishing seven papers and you are only publishing four,' and one of those four is the discovery of RNA interference, that clearly would be the wrong way to think about things," he says.

Raphael Kopan, a developmental biologist and NIGMS grantee who this year ran his lab at Washington University in St Louis on \$800,000, says that Berg should be applauded for trying to scientifically analyse what his institute gets

for its investment. But without segregating the data — comparing, for instance, investigator-initiated grants with projects instigated by the NIGMS, or intramural with extramural investigators — "it may lead to the wrong conclusion — that scientists do best if their funds are limited and their labs are small. I don't think this is necessarily correct," says Kopan.

Still, Berg's analysis has served a purpose: validating a 20-year-old NIGMS policy of generally denying new grants to well funded labs. Since 1999, that has meant labs with more than \$750,000 in direct support from all sources, including the award being applied for.

Marburger says that Berg's analysis provides a "reality check" of that policy. The results, he says, are "an indication that they aren't making a big mistake".

Berg's next project will be to tackle the impact of the abbreviated grant-application forms that came into effect at the NIH in January. Among other things, he will be asking whether and how the slimmed-down form for the agency's mainstay grants is affecting the scores that applicants receive.

Whatever happens, the future is likely to bring more austerity, making it important for defenders of science agencies to arm themselves with the best quantitative ammunition they can generate. In this environment, questions such as Berg's "are very good to ask", says Kopan, who argues that Congress is already effectively cutting the NIH by failing to keep its budget growing as quickly as the costs of doing biomedical research. If cuts have to be made, he says, "we might as well go ahead and do it correctly". ■

➔ **NATURE.COM**  
Are measures of  
scientific productivity  
fair? Visit:  
[go.nature.com/nj2xqk](http://go.nature.com/nj2xqk)

## FUNDING

# UK science will be judged on impact

*Pilot scheme paves way for university research to be awarded on the basis of society benefits.*

BY NATASHA GILBERT

Research funding agencies have long dreamed of favouring scientists who have a track record of turning their work into tangible benefits for society and the economy. Attempts to judge 'impact' have been controversial, but the UK government thinks it has hit on a workable scheme. Last week, the Higher Education Funding Council for England (HEFCE) unveiled the results of a year-long pilot study that showed that using peer-review panels to assess the impact of research in UK universities is "workable" and "robust".

The idea of getting tangible returns from research funding aligns with the current coalition government's demands that researchers "do more for less", in the words of business secretary Vince Cable. With the success of the pilot study, the method looks set to become a key part of the nation's research audit system by 2014. This Research Excellence Framework (REF) will replace the Research Assessment Exercise (RAE), which did not factor research impact into its calculations, and will be used to apportion more than £1.5 billion (US\$2.4 billion) per year. Research impact is expected to contribute up to 25% to the overall rating of a university department's research quality.

In the pilot study, university departments submitted case studies describing the impact of the work done by one in ten of their researchers over the past 17 years. Other academics and industry scientists on subject-specific panels reviewed the case studies, and awarded rankings from 4\* (the best) to unclassified. Eleven

University projects with clear advantages for society, such as bumblebee conservation, will be cited to win funding.

physics departments and ten departments of clinical medicine and of Earth systems and environmental science took part in the exercise. 'Impacts' included the establishment of spin-out companies, influence on policy relating to the environment, or the development of products such as computer software or technology.

Many academics are concerned that the added focus on research impact would skew funding towards applied research. Jonathan Grant, president of RAND Europe, a research consultancy based in Cambridge, UK, wrote a report last year criticizing the REF, and argues that impact should determine only 10–20% of universities' funding to avoid channelling funds away from blue-skies research. However, the pilot's successful use of peer-review panels has convinced him that "if

you are going to measure impact, this is the way to do it".

HEFCE will unveil a final plan for the REF in February 2011, but universities say there are still some problems to be ironed out. Anna Grey, research manager at the University of York, UK, says that some of her university's industry partners were not happy to release the details it needed to demonstrate impact, such as financial savings made as a result of products developed by the university. "Unless we can prove to the companies that the information will remain confidential, we will struggle to get hard evidence of impact," she says.

And Peter Main, director of education and science at the Institute of Physics in London, worries that universities could pressure departments to continue research in fields that have generated impact in the past, "even when more future impact might be generated from new directions". ■



IMAGEBROKER/FLPA

➔ **NATURE.COM**  
For a longer version  
of this story, see:  
[go.nature.com/qbkcgf](http://go.nature.com/qbkcgf)



The German Centre for Neurodegenerative Disease will be based at a hub (above) to be built in Bonn.

MEDICAL RESEARCH

# Germany plans for healthy future

*National health-research centres take shape.*

BY ALISON ABBOTT

After years of wrangling, Germany is finally setting up a national medical-research system, meant to help the country's scientists compete with powerhouses such as Britain and the United States.

German biomedical research has traditionally been done by universities and their clinics (which have direct access to patients), and by research institutes belonging to organizations such as the Max Planck Society and the Helmholtz Society. But they don't always collaborate, limiting the effectiveness of the country's research in this field.

Three years ago, federal research minister Annette Schavan decided to create a series of national health-research centres that would bring together these disparate efforts to make more efficient use of funding, enable multidisciplinary studies in translational medicine and attract top talent from abroad. The first such centre — the German Centre for Neurodegenerative Disease (DZNE), headquartered in Bonn but with seven partner institutes around the country — opened last year without controversy. But angry protests greeted proposals for a second, for diabetes research. Some university medical faculties claimed that power was being given to research centres lacking appropriate expertise, and the widely publicized row threatened to derail plans for subsequent centres.

Those plans are now firmly back on track. The National Centre for Diabetes Research (DZD) opened officially on 9 November; the

day before, the government approved the locations of four other distributed national centres, covering cardiovascular diseases, infectious diseases, lung diseases and cancer. These should begin operation next year.

All the new centres will focus on translational medical research and each will receive around €35 million (US\$48 million) a year in federal government funding (the DZNE receives more). To defuse tensions with the universities, the ministry is now letting participating scientists decide how to structure the four new centres — unfamiliar territory for many of them. "We don't yet know how to do this," confesses immunologist Dirk Busch at the Technical University of Munich, a member of the infectious-diseases centre.

To circumvent an existing bar on federal funding for universities, federal money for each centre will be funnelled through a research institute of the Helmholtz Society, which is 90% federally funded. Most of the new centres have been chosen to give equal importance to each of their half a dozen or so bases, which were selected by international expert committees from competing bids. Each bid was required to be a collaboration between local universities and non-university research institutes. Once established, the centres may draw in research groups not based in the winning locations. The winners now have just a couple of months to put together a concept for the centres' organization and research, to be submitted to the same review committees. The ministry wants money to flow before the end of 2011.

## TECHNOLOGY

# No rest for the bio-wikis

*Biologists' collaborative data repositories come of age.*

BY EWEN CALLAWAY

Most Tuesdays, a group of scientists at the Wellcome Trust Sanger Institute in Hinxton, UK, meets over lunch to edit Wikipedia pages. But there is no obsessing over the minutiae of Britney Spears's career to be found here — instead, they are building the next generation of global biological databases.

"Yesterday, we created 18 microRNA articles," says Alex Bateman, a computational biologist at the Sanger Institute who helped to found Rfam, a database that includes a set of Wikipedia articles covering about 1,500 families of RNA molecules, maintained by more than 2,000 editors. In the past few years, community-curated biological websites such as this have multiplied, and on 29 November scientists will gather for a first-of-its kind conference called Biological Wikis, in Naples, Italy, to take stock of the 'bio-wiki' approach and plan future expansion.

Wikis, collaboratively edited web pages named after the Hawaiian word for quick, offer a solution to the growing data glut in biology. Conventional databases are struggling to keep up with the flood of information about genes and proteins that labs are amassing by the terabyte ( $10^{12}$  bytes). "The old model of annotation, where the central database handles that information, doesn't work," says Dan Bolser, a computational biologist at the University of Dundee, UK, who is involved in curating a protein-structure database called PDBWiki.

So biologists increasingly maintain and update web pages focused on particular genes or proteins — or any concept or object of interest (see 'The bio-wiki boom'). The success of Wikipedia, the ubiquitous online encyclopaedia compiled using the same technique, proves that community annotation works, say bio-wiki enthusiasts. "There's a genuine generational and

technological change that's happening," says Ewan Birney, a bioinformatician at the European Bioinformatics Institute in Hinxton.

At the Naples meeting, scientists will discuss the lessons of the handful of bio-wikis that are beginning to assemble a critical mass of readers and contributors. These bio-wikis are now attracting more contributions from the community than from the developers themselves, and advocates say that the sites are becoming indispensable tools in some areas of biology.

Gene Wiki, for instance, has more than 10,000 Wikipedia pages, each devoted to a single gene, and draws some 4 million views and 1,000 edits every month. Many scientists come to the site after an experiment identifies a laundry list of genes that are of interest in their work — and which they know little about or have never heard of, says Andrew Su, a bioinformatician at the Genomics Institute of the Novartis Research Foundation in San Diego, California, and one of the driving forces behind Gene Wiki. "It's a great way to go in and get up to speed," he says.

Bio-wikis that are hosted by Wikipedia benefit from the contributions of its existing altruistic community of 'Wikipedians', says Bateman. His team will soon launch a protein-family wiki that will also be hosted on Wikipedia. "One of the big surprises for me in all of this is that the contributions we're getting are as much from non-scientists as they are from scientists," he says. Non-specialists with enough interest and expertise to add information about the activity of microRNAs may be rare, but others can provide help with page formatting and standardization, which are "important, valid contributions", says Bateman.

But Wikipedia comes with its own rules and idiosyncrasies, which limit its usefulness for some kinds of biological data. To merit a page on Wikipedia, a subject — whether a gene, ►

ENGEL, UKB

After more than a year of uncertainty, the mood is upbeat. "Germany has not been competitive in health research in the past decades," says Oliver Eickelberg of the Comprehensive Pneumology Centre in Munich, a member of the new lung-research centre. Even sceptical universities are coming round to the idea. Clinician-researcher Andreas Zeiher, from the Goethe University in Frankfurt, who is a member of the cardiovascular-research centre, says they now realize "that it is better to have something than nothing".

One project that the lung centre is discussing is the identification of molecular signatures of lung fibrosis in order to develop targeted therapies. A single clinical centre would typically have only some 150 such cases to work with, but by pooling patients from around Germany the lung centre expects to be able to recruit up to 1,000. This number would make analysis — using the sequencing and systems-biology platforms the centre intends to establish — statistically feasible. "The new centres allow us to concentrate our activities with the security of long-term investment and increase our visibility internationally," says Eickelberg. "That's important if we want to attract outside investigators."

"Germany is making a smart move," says David Warburton, a physician-scientist at the Children's Hospital Los Angeles at the University of Southern California, and a member of the expert advisory panel for the lung-disease centre. He believes that synergies between groups involved with the new centres will add value to the "well-funded, well-organized but decentralized research activities in Germany".

Participants agree that the effort will succeed only if it is backed by stable funding, and in a statement to *Nature* the ministry confirmed that this is its intention. Oncologist Otmar Wiestler, head of the German Cancer Research Centre in Heidelberg, is confident that the government will not allow the centres to fail. "This is a new era for biomedical research in Germany — nothing less." ■



## MORE ONLINE

### STAY CURRENT

- Demonic device converts information into energy [go.nature.com/g9oqpr](http://go.nature.com/g9oqpr)
- ITER fusion project begins hunt for budgetary savings [go.nature.com/djzau8](http://go.nature.com/djzau8)
- Beating heart cells controlled by light [go.nature.com/zaxrhz](http://go.nature.com/zaxrhz)

### VIDEO



Caught on camera: the secret of how cats lap milk [go.nature.com/8ka2wt](http://go.nature.com/8ka2wt)

IDA MINU/LAWRY



► a protein or any other biological entity — must be considered noteworthy by the Wikipedia community. Important data, such as protein crystal structures and genetic variants, do not always qualify, says Su.

The rub is that many bio-wikis not housed within Wikipedia struggle to attract readers and editors. But Alexander Pico, a bioinformatician at the Gladstone Institute of Cardiovascular Disease in San Francisco, California, thinks that these problems will fix themselves. “The vision going forward is that more and more scientists will be involved in the curation and consumption of data and they won’t need to accidentally stumble on it through Wikipedia,” he argues. His team’s WikiPathways site, which characterizes and visualizes biological pathways and is independent of Wikipedia, thrives because the systems biologists it attracts are already avid consumers of other people’s data and therefore see the benefits of a wiki, says Pico.

One challenge to bio-wikis that will be addressed at the Naples meeting is their text-based default layout, says Su. Written entries devoted to individual genes and proteins fit well within Wikipedia. But the format is a poor match to the highly structured, searchable data

## THE BIO-WIKI BOOM

Collaboratively edited biological databases help the community keep up with a flood of information.

Wiki	Topic	URL	Pages
EcoliWiki	Genes and proteins in <i>Escherichia coli</i>	ecoliwiki.net	63,784
PDBWiki	Protein structures	pdwiki.org	64,071
Wikigenes	Genes, proteins and biochemicals	www.wikigenes.org	123,332
Rfam wiki	RNA families	en.wikipedia.org/wiki/User:Ppgardne/Rfam10.0	824
Gene Wiki	Genes	en.wikipedia.org/wiki/Portal:Gene_Wiki	10,118

sets favoured by computational biologists, which include the precise relationships between genes, proteins and other factors. A number of bio-wikis, including Su’s Gene Wiki, are adopting a software package called Semantic MediaWiki. This will bring them closer to working like true databases: for instance, the software could allow scientists to search for all the proteins phosphorylated by a specific kinase enzyme expressed in a particular tissue, rather than having to look up each interaction individually.

Despite such innovations, bio-wikis might not truly take off until scientists can get career-advancing credit for contributing to them.

“Editing your wiki is not going to get you your grant, it’s not going to get you promoted,” says Jim Hu, a molecular biologist at Texas A&M University in College Station and one of the founders of EcoliWiki, a repository of information about the model bacterium *Escherichia coli*. One database trying to solve the attribution problem is Wikigenes, a site devoted to annotating 120,000 genes and other biomedical concepts, which meticulously records and displays the individual contributions of its 1,800 active editors.

Persuading funding agencies and tenure committees to take those contributions seriously would mark a major milestone for bio-wikis. Until then, says Bolser, “it’s not clear to scientists why they should spend time editing a wiki article if it just gets them kudos from a few geeks on Wikipedia”. ■

### CORRECTION

The Editorial ‘A painful remedy’ (*Nature* **468**, 6; 2010) misspelt the name of physicist Jan Hendrik Schön as ‘Hendrick’ and incorrectly gave his nationality as Austrian. He was born in Germany.



# THE HOTTEST YEAR

The release of climate-science e-mails last November ripped apart Phil Jones's life. He's now trying to patch it back together.

BY DAVID ADAM

**I** like to think the worst is over, but it's coming up to the first anniversary and it's something I'll always remember at this time of year, when the nights close in. This is the time it happened."

Twelve months ago, Phil Jones was a productive, if not particularly outspoken, climate scientist. That was the way he liked it. Head of the Climatic Research Unit (CRU) at the University of East Anglia (UEA), UK, Jones worked with the Met Office to compile data from weather stations around the world into a monthly series showing global average temperature. He had much on his mind — not least

a puzzling drop in North Atlantic sea surface temperatures during the mid-twentieth century that he had recently helped to discover. It was a curious finding, but Jones would soon have bigger things to ponder.

On 19 November 2009, someone released roughly 1,000 e-mail messages and documents stolen from a server at the CRU. Many of them contained Jones's private correspondence, which sometimes showed him in an unflattering light.

He gloated about the death of a prominent

climate sceptic, and suggested to colleagues they should delete e-mails to keep sceptics from gaining access to information. Most famously, he boasted that he had used a "trick" to "hide the decline" in a temperature chart.

Very soon, members of the sceptic community had pounced on these messages as evidence that Jones and others had concealed flaws in their temperature data and abused the peer-review system to gag critics of climate researchers. Jones faced a storm of accusations that ranged from scientific misconduct to plans to install an autocratic world government through the spread of false hysteria about

global warming. He received some 200 abusive or threatening e-mails, the most troubling of which targeted him and his family. "Someone, somewhere, will hunt you down," read one. "You are now blacklisted," read another. "Expect us at your door to say hello."

The e-mails also triggered several official investigations, including one by the UK Parliament, which ultimately determined that Jones had not committed any serious offences. Case closed.

Not for Jones, who still faces attacks from critics and is trying to cope with unwanted memories as the anniversary approaches (see a 'Career by degrees'). Never comfortable with the media, Jones has given few interviews since the controversy began. But as part of an attempt to put the past year behind him, he agreed to show *Nature* around the CRU earlier this month and to talk at length about his experience. He proved largely unrepentant.

Aged 58, Jones looks far better than during the darkest days of last winter, when he was spiralling downhill and even contemplated suicide. Colleagues were stunned by his decline.

## "I did wonder why supportive scientists didn't go to the media."

Jones was never an extrovert, but he withdrew further and his mental collapse was mirrored by a rapid loss of weight.

In March, when a frail and hesitant Jones answered questions before an investigating parliamentary committee, his appearance reminded many of the distressing 2003 case of David Kelly. Kelly was the UK weapons inspector outed as the source of a media story about government exaggeration of Iraqi weapons of mass destruction. He was also questioned by a parliamentary committee — and subsequently killed himself. "I made the connection," Jones says about the Kelly case. "But I didn't talk about it."

Jones has regained much of the lost weight, and he no longer takes the medications that kept him calm during the day and asleep at night. He is back in charge of the CRU (he stood aside for some eight months while enquiries were pending). So, how have events of the past 12 months changed him?

"I'm a little more guarded about what I say in e-mails now," he says. "One thing in particular I'm doing is not responding so quickly. I might have got an e-mail in the past and responded with an instant thought in the next 10 to 15 minutes, whereas now I might leave it a day."

Jones admitted in the parliamentary inquiry to sending some "awful e-mails", but defends the right of scientists to express themselves in what they consider personal communications. "People would be saying much the same things at scientific meetings and discussed [them] over dinner. But in an e-mail, it is recorded. People have probably forgotten what you said after a night out."

Although other scientists were quick to defend the reality of man-made global warming, public support for Jones was harder to find. Officially, senior figures in the UK science establishment say this was because they did not want to prejudice ongoing enquiries. Privately, they say that the e-mails looked bad, and should the CRU scientists have been found guilty of misconduct, they did not want to get dragged down with them.

"I was getting lots of messages of support from my fellow scientists," Jones says. "And I did wonder why they didn't go to the media and say the same things they were saying to me."

The CRU server that held the stolen information was seized long ago as evidence from the cluttered desk where it sat in one of the unit's cramped offices. The unit itself is housed in a curious four-storey cylindrical tower at the heart of the busy UEA campus, and it brings to mind a Norman keep within a medieval castle. An appropriate analogy, considering that its occupants have weathered an extended siege that left visible scars on the tower's exterior. Its

doorbell was removed to shield the scientists inside from the incessant ringing of journalists and film crews.

Outsiders are often surprised at how small the unit is, with just three full-time staff scientists. Jones's office is on the top floor, where the computer on which he typed many of the e-mails sits amid a carpet of scientific reports and papers. Keith Briffa, a tree-ring specialist, has an office across the landing. Climate researcher Tim Osborn is next door, struggling with a familiar problem. "My inbox is full and I need to delete some e-mails." Then, with a thin smile: "But I'm not allowed to now, am I?"

Temperature data analysed by these researchers serve as the foundation for countless studies, which have steadily identified and analysed the signal of global warming caused by human activities. The growing importance of this work made Jones and other CRU scientists a target for Internet bloggers sceptical of their methods and the conclusions drawn from them. Long before the e-mail scandal, Jones and his team found themselves fielding enquiries about their research from outside the conventional scientific community.

An independent inquiry headed by former senior civil servant Alastair Muir Russell examined many aspects of the work done at the CRU, looking specifically to see if the centre had committed fraud or some other type of scientific misbehaviour. The investigation found no reason to doubt the honesty and integrity of the CRU scientists, but it did criticize the way those scientists

responded to information requests, or in some cases, failed to respond. The report said there had been a "consistent pattern of failing to display the proper degree of openness".

Some scientists echo these conclusions. Mike Hulme, a climate researcher at the UEA who worked at the CRU from 1988 to 2000, said that certain aspects of the culture in the research unit were "unwise and unhealthy". He notes in particular that the CRU was slow and inconsistent in responding to data requests, and says it suffered from "intense tribalism". But Hulme says the work at the CRU "was not fraudulent, and certainly did not justify the personalization of the attacks subsequently made on them".

In his defence, Jones says he wrestled with how open scientists should be to requests for information. "I started responding to those back in 2003 and 2004, but they just asked more and more questions and it was just a drain on resources. That's when things probably went awry." He claims he changed tack when he saw that the information he supplied was not used by those who demanded it. Rather, each response

simply triggered more questions. "I just realized it was taking up too much time," he says.

By failing to answer all requests properly, Jones says he wasn't acting any differently from other researchers. "There are some people I have sent requests to, other scientists, who have never replied. I've asked people for data and reprints of papers and I've never got a response. So I think I responded quite well and the CRU responded quite well."

Jones complains frequently about distractions from his research. "The amount of time we get to do research just seems to be less and less, and you see things that take away that research time, or you find yourself working at weekends or in the evenings to the annoyance of your family." Autumn is a "bad time" because his teaching load increases. He got frustrated with meetings with university officials to discuss freedom of information requests because "it takes away your research time". And he rarely agrees to peer review scientific papers. "If you start doing lots of reviews, you find that your quality research time also goes."

When he did review papers, the stolen e-mails revealed, he told colleagues he "went to town" to make sure that those manuscripts he did not like were not published. The Muir Russell report found there was no abuse of peer review and said such robust exchanges were typical in science. Jones says he learned long ago that he needed to be absolutely clear with editors, because in the past he had written what he thought were critical reviews only to see the papers in question get published. "I realized that to make sure an editor rejects a paper you have to go a bit stronger in the review."

**NATURE.COM**  
Comment on this  
story at  
[go.nature.com/dg9cmj](http://go.nature.com/dg9cmj)



## A CAREER BY DEGREES

**1976** Phil Jones joins the Climatic Research Unit (CRU) at the University of East Anglia, UK, where he will spend his entire career to date.

**1982** The CRU publishes its first monthly global average temperature series based on weather-station data (below).



**1990** Jones co-authors an influential *Nature* paper that shows urbanization is not responsible for increasing temperatures.

**1999** Jones e-mails colleagues, saying he used a "trick" to "hide the decline" in preparing a temperature chart. The decline refers to late-twentieth century tree-ring data that suggested a cooling, in contrast to the temperature data.



**2005** Britain introduces the Freedom of Information (FOI) Act, giving critics a legal route to demand data from Jones and the CRU (above).

**July 2009** The CRU receives 58 FOI requests in under a week as part of a blog campaign.

**November 2009** Some 1,000 e-mails and documents stolen from the CRU are released on the Internet.

**February 2010** Jones tells *The Sunday Times* he considered killing himself after the e-mails were released. He subsequently receives e-mails telling him to do so.



**March 2010** Jones appears before an inquiry (above) by a parliamentary committee on science and technology.

**July 2010** The Muir Russell inquiry into the CRU e-mails clears scientists of serious charges, but criticizes their response to FOI enquiries.

**November 2010** Jones tells *Nature* he is on the mend, but still fears more e-mails could be released in the future.

He adds: "The whole point about trying to pervert the peer-review process is that it is impossible to do it. There are so many journals and if people are persistent enough, they can get their papers published."

Another allegation was over his use of data from weather stations in China for a 1990 paper on the impact of urbanization on temperature. The paper<sup>1</sup>, published in *Nature*, stated that data were used from stations where there had been few, if any, changes in instrumentation, location or observation times. When critics later uncovered the fact that many of the stations had moved, they cried fraud; earlier this year, Jones said in a separate interview with *Nature*<sup>2</sup> that he was considering a correction.

He now says such a step is unnecessary and that he stands by the claims in the paper. He was on medication during the previous inter-

**"I don't know that I can offer advice.  
Whatever you try to do, the goalposts keep moving."**

view, he says, and felt under pressure then to publicly concede that he had made mistakes.

He says the description of weather-station movement "has been completely misinterpreted". The set of 84 Chinese stations referred to in the paper were drawn from a larger group of 265, for which the Chinese had location histories. Jones and his colleagues did not claim that none of the selected stations had moved, only that they picked out ones that had moved the least, he says.

Such shifts do not significantly affect results, Jones says, because there was no general pattern to the station relocation: on average, ones moving to colder places were balanced by ones moving to warmer spots. But the Chinese scientist who supplied the station information has now retired and the authorities there have not released the full station-history data — making it impossible for Jones, he says, to provide the evidence to support the statement.

One issue critics continue to badger Jones about is whether he deleted e-mails that had been requested through the freedom of information process. Jones insists he never did, as that would have qualified as an offence. What about deleting e-mails that could be requested by future freedom of information requests? Britain's Information Commissioner's Office, which adjudicates such cases, says it is allowed. However, the Muir Russell report said that this kind of pre-emptive deletion is not consistent with the "spirit and intent" of the law, and there is evidence that CRU scientists took that questionable approach. When Jones is now asked if he deleted such messages, he says: "No, I deleted e-mails as a matter of course just to keep them under control."

So why did he urge colleagues to delete messages in which they discussed, among other things, the preparation of a report for the

Intergovernmental Panel on Climate Change? An attempt to thwart critics, perhaps? "That was probably just bravado at the time," he says. "We just thought if they're going to ask for more, we might as well not have them."

Then Muir Russell was correct? Had Jones broken the spirit of the law? "Not necessarily, if you've deleted them ahead of time," he says. "You can't second guess what's going to be requested." Jones goes back and forth on his motivations. Deleting e-mails would simplify his life if people requested them in the future, but that was not why he got rid of them, he says. "I deleted them based on their dates. It was to keep the e-mails under control," he repeats.

A source close to the CRU says it is almost impossible to determine who deleted what and when — much less why. More certain is the conclusion that the hack of the server was

a sophisticated attack. Although the police and the university say only that the investigation is continuing, *Nature* understands that evidence has emerged effectively ruling out a leak from inside the CRU, as some have claimed. And other climate-research organizations are believed to have told police that their systems survived hack attempts at the same time.

Jones and others connected to the CRU fear the hackers may be sitting on more stolen e-mails, but Jones feels confident the worst is behind him. "It really is not somewhere I would like to go through again. But having been through it once, I think I am a bit hardened to it."

Can Jones offer any advice to research scientists who wake up one morning to find themselves the centre of a worldwide scientific scandal? "I don't know that I can. The thing to point out is that whatever you try to do, the goalposts keep moving." As soon as he responded to one criticism, another popped up.

Jones has steadily begun to piece together his professional as well as his personal life. The discovery of the sudden Atlantic cooling was recently published in *Nature*<sup>3</sup> and he has started to attend conferences again. He agrees to pose for photographs outside the CRU building, gazing at the blue sky. Then he shuffles back into the relative calm of his unit: one scientist who now realizes his castle walls cannot completely shield him from the outside world. ■ **SEE EDITORIAL P.345**

**David Adam** is an editor with *Nature* in London.

1. Jones, P. D. et al. *Nature* **347**, 169–172 (1990).
2. Heffernan, O. *Nature* doi:10.1038/news.2010.71 (2010).
3. Thompson, D. W. J., Wallace, J. M., Kennedy, J. J. & Jones, P. D. *Nature* **467**, 444–447 (2010).



# Scientist as star

Sleep researcher Sara Mednick has straddled the line between media darling and respected scientist. But why is there still a line at all?

BY ERIK VANCE

Sara Mednick was flying high in January 2007. She was doing a television appearance a day, every day, for a month. And she was being featured on radio shows around the United States, repeating talking points from her just-released book, *Take a Nap! Change Your Life*. Dozens of businesses were calling for her expertise and endorsements, including the Silicon Valley juggernaut, Google, which requested a 'napping strategy' for its employees. By all appearances, Mednick had joined a class of scientists that spans academia and popular culture with aplomb.

But it wasn't easy. "It's such a crazy experience where you are in a different city every day, and you're working these ridiculous hours to do these daybreak TV shows," she says. She was baffled by the experience, and a little flattered. "There was a part of me that was wondering, could I still do my work and try to also be this next big thing?"

It is a question being asked by a rising number of scientists, as the 24-hour news cycle and proliferation of media outlets and blogs have made achieving 15 minutes, or more, of fame easier than ever. Polls suggest that the scientific community want a better portrayal of science in the media, but are unsure whether they should be the ones to provide it. A 2009 study by the Pew Research Center in Washington DC found that 85% of scientists see the public's lack of scientific understanding as a major problem, and most were unimpressed with the traditional media coverage of the subject. Still, a poll by *Nature* earlier this year suggests that many researchers think that their institutions put little emphasis on press exposure and that it shouldn't be a major factor when determining career advancement (see [go.nature.com/em7auj](http://go.nature.com/em7auj)).

That is a tide that is changing, says Stephen Hinshaw, a psychology department chair at the University of California, Berkeley. What might have been seen by previous generations as garish or vain is quickly becoming another part of a scientist's workday. "Years ago, somebody who was media savvy would have been viewed pejoratively as too slick. Today, it could well be an advantage, given fundraising, appealing to donors and appealing to a wide audience to make psychological science relevant. All of those are good things."

But as Mednick's story shows, celebrity science is

**NATURE.COM**  
See videos of  
Sara Mednick at:  
[go.nature.com/zb8kk4](http://go.nature.com/zb8kk4)



not all good. She has had an impact on people outside the tight-knit circle of her scientific peers and enjoyed the celebrity status. But she still wants to be seen as a serious scientist in traditional academia. She has found that scientific celebrity needs to be maintained, rarely pays and can have unintended consequences on one's professional reputation.

Mednick conducts her research at a sleep laboratory at the University of California, San Diego. "We need to be really quiet," she says, gently closing the door to her office. "Someone is napping in the next room."

The lab consists of hotel-like rooms for napping, plus rooms for researchers to monitor sleeping subjects — quietly. Despite being there for five years, keeping quiet still seems to be a struggle for Mednick, who has piercing blue eyes and an eruptive laugh. Within a few minutes, she seems to have forgotten the person sleeping in the next room and is animatedly describing her work.

Colleagues refer to Mednick as one of the world's leading experts on naps. Her work looks at various types of sleep and its effect on human cognitive and motor skills. She and her colleagues have shown, for example, that 60- and 90-minute naps can improve performance as much as a full night's sleep on several visual-perception tasks (S. Mednick *et al.* *Nature Neurosci.* 6, 697–698; 2003).

Mednick's fascination with naps started in the late 1990s when she was a psychology PhD student at Harvard University in Cambridge, Massachusetts, studying visual memory in patients with schizophrenia. But after hearing lectures by sleep expert Robert Stickgold, she decided she wanted a new direction. She started working with Stickgold at Harvard, and later landed a postdoc position at the Salk Institute in San Diego, California, in 2003. In the competitive academic atmosphere at Salk, colleagues expected her to write as many papers as possible and then go on to a tenure-track position.

Instead, Mednick spent her final postdoc year writing a book on napping for the public. Her publisher, Workman Publishing, is a New York company that prints titles such as *The Cake Mix Doctor*, *How to Satisfy a Woman Every Time* and *The Betty White Wall Calendar*.

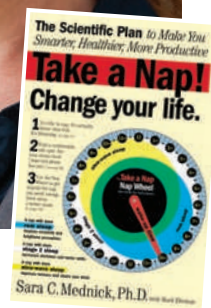
"What the hell are you doing?" That's what all my scientific friends were saying," she says. "'This is not helping you get tenure.'" Mednick says that she wrote the book, together with co-author Mark Ehrman, because she wanted her research to reach people. "It was such an obvious book to write," she says. "I just like the idea of having my research being real world." She concedes that vanity and the hope for a pay cheque were a small part of the motivation. Ultimately, however, Mednick seems driven by a desire to overturn conventions. A former actress, Mednick marches to her own drumbeat, say friends and colleagues. The book definitely got her noticed — leading to the whirlwind of media attention in 2007.

Widespread preoccupation with sleep science has fostered a bustling book market. Amazon.com carries more than 750 titles under the headings 'sleep' and 'medicine'. Only a minority of these have been written by scientists with experience in sleep research (about one-third of the 30 top-selling authors have advanced degrees). Many of the rest are written by self-help gurus, yoga teachers and even pastors. So the media jumped at the chance to talk to Mednick: a bona fide scientist with evidence that midday naps were beneficial.

Despite some 150 media appearances and countless interviews, however, Mednick's book only netted her about US\$30,000, which barely covered her advance. She says that Google did not pay her for the consulting work she did. A Google representative said the company could



Sara Mednick's book proselytizes the public to the power of napping.



not provide details of the arrangement. The only corporate money she received was from the Dutch company MetroNaps, which markets a futuristic napping 'pod' for snoozing at work. Mednick says she made \$10,000–\$15,000 designing sleep survey questions for the company's website, and to this day has been unable to convince them to remove her picture.

"It was before I really knew what I was doing," she says. "I allowed them to use my picture and my name. I suddenly realized that that wasn't at all what I wanted to be affiliated with."

Back in her lab, Mednick goes into the monitoring room, fretting for a moment that the noise in her office has disturbed the subject. Her current study is examining the benefits of short bursts of rapid-eye-movement (REM) sleep, so she needs the nappers to sleep well. According to an electroencephalography readout — which records the electrical activity of the brain — this individual has had a fitful nap.

Much of Mednick's research, as well as her book, looks at the best nap length and

the best time of day to take one. To illustrate this, she and Ehrman have designed a 'nap wheel' to help people to visualize their sleep schedule. But nap wheels don't exactly further one's career. Mednick has won grant money for her research but is still looking for a tenure-track position. "She is taking a risk," says James Maas, creator of several educational documentaries on sleep and author of the *New York Times* bestselling book *Powersleep* (Harper, 1998). "I would have advised her to wait until she had tenure," says Maas. He says that few academics would openly criticize such behaviour but that it can affect scientists more subtly, tarnishing them in the eyes of funders, for example, who question the dedication to daytime TV shows rather than the lab (for more on the rewards and potential pitfalls of media engagement, see page 465).

Stickgold says that Mednick's public persona has undoubtedly affected her career, but in ways that are hard to spot — a missed grant opportunity or a keynote address being offered to someone else, for example. Mednick can't point to specific instances in which this has happened. She does lament the fact that she has not managed to publish in either of the field's primary journals, *Sleep* and the *Journal of Sleep Research*, even though she has published in higher-impact mainstream journals.

David Dinges, editor-in-chief of *Sleep*, says that Mednick is "a respected scientist who has done interesting work," but that 75% of all submitted manuscripts are rejected. Mednick doesn't blame the journal, but is concerned that her outside activities could hinder her progress. Even so, she claims to have no regrets about her book or media presence. She continues to make television appearances and write for the popular press. And she advises younger colleagues to do the same.

Mednick is still deciding where she belongs. But every step in the direction of celebrity has to be negotiated carefully. In late August, Mednick got a call from the popular talk show, *Dr. Phil*, known for high-drama confrontations. The talk-show producers said they loved her book and were interested in making a show about sleep. In the end, however, they decided to avoid what they called 'the scientific route', instead opting for someone to interpret the dreams of women who think their partners might be unfaithful.

"Probably for the best," says Mednick. ■ [SEE CAREERS P465](#)

Erik Vance is a freelance writer in the San Francisco Bay Area, California.



# COMMENT

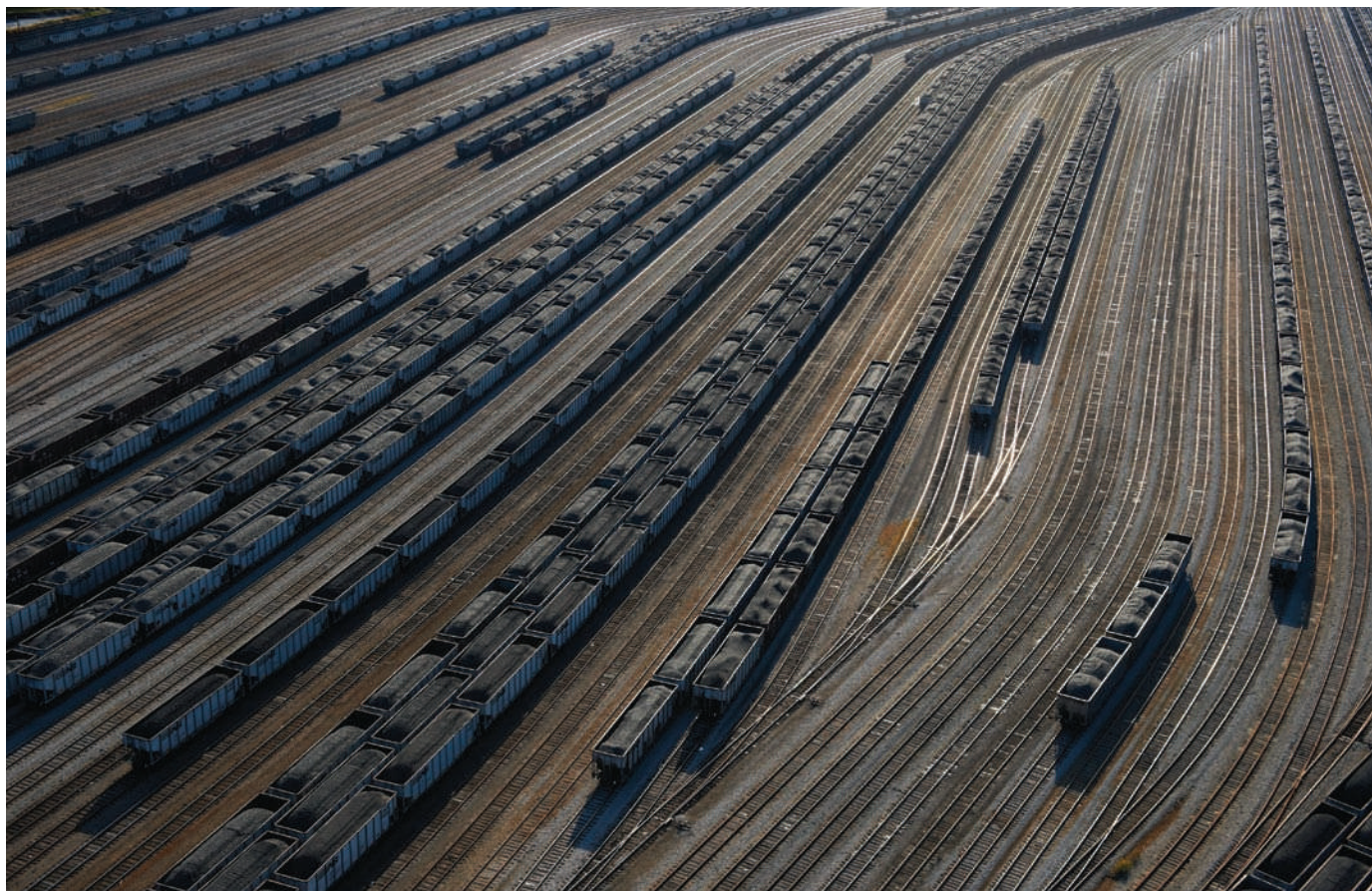


**TECHNOLOGY** Evolution of mechanical systems moves beyond human control **p.372**

**THEATRE** Rosalind Franklin drama captures the zeal and the back-stabbing **p.375**

**BIODIVERSITY** Amateur taxonomists describe many new species in Europe **p.377**

**OBITUARY** Benoît Mandelbrot, father of fractal geometry, remembered **p.378**



C. DAVIDSON/CORBIS

As coal reserves are depleted, busy coal-train facilities, such as this one in Norfolk, Virginia, will become a thing of the past.

## The end of cheap coal

New forecasts suggest that coal reserves will run out faster than many believe. Energy policies relying on cheap coal have no future, say **Richard Heinberg** and **David Fridley**.

World energy policy is gripped by a fallacy — the idea that coal is destined to stay cheap for decades to come. This assumption supports investment in ‘clean-coal’ technology and trumps serious efforts to increase energy conservation and develop alternative energy sources. It is an important enough assumption about our energy future that it demands closer examination.

There are two reasons to believe that coal prices are likely to soar in the years ahead.

First, a spate of recent studies<sup>1–5</sup> suggests that available, useful coal may be less abundant than has been assumed — indeed that the peak of world coal production may be only years away. One pessimistic study<sup>1</sup> published in 2010 concluded that global energy derived from coal could peak as early as 2011.

Second, global demand is growing rapidly, largely driven by China. Demand rose modestly in the 1990s (0.45% per year), but since 2000 it has been surging at 3.8% per year. China is both the world’s biggest producer of

coal (40% of global production) and its biggest consumer. Its influence on future coal prices should not be underestimated.

Economic shocks from rising coal prices will be felt by every sector of society. Better data on global coal supplies is long overdue and energy policies that assume a bottomless coal pit need rethinking urgently.

Forecasting future supplies of coal is a murky business, largely because of the unreliability of national estimates. China claims that it has enough coal to fuel its ▶



▶ growing economy at current rates. According to data collected in the 2000–10 national resource survey by the China's Ministry of Land and Resources, the country's proven reserves of coal total 187 billion tonnes, the second-largest reserves after the United States. For China, that is about 62 years' worth of coal — at 2009 rates of consumption (roughly 3 billion tonnes a year). This simple 'lifetime' calculation is popular with industry and politicians but it can generate a false sense of security over the actual state of reserves.

'Proven recoverable reserves' are estimates of the national coal resources that geologists believe are technically and economically feasible to mine. New mining technology and higher coal prices could, in principle, increase the size of those reserves. But the overwhelming global trend, as revealed by national coal surveys over the past few decades, is for the size of countries' estimated reserves to shrink as geologists uncover restrictions — such as location, depth, seam thickness and quality — on the coal that can be practically extracted.

For example, both German and South African reserves have fallen by more than one-third between 2003 and 2008. The first British coal survey, in the nineteenth century, suggested that the nation had enough coal to last 900 years. The current reserves lifetime is only 12 years<sup>6</sup>, and the British coal industry is a tiny fraction of its former size. Similarly, the first official US coal survey, in the early twentieth century, suggested that the country had enough coal for 5,000 years. That estimate shrank to about 400 years in 1974 and stands at 240 years today. There are exceptions to this trend: estimates of reserves in Indonesia and India have grown. However, in aggregate, estimates of global coal reserves have dropped at a faster rate in recent years than can be accounted for by mining alone.

### OPTIMISTIC FORECASTS

China's reserves were last surveyed in the early 2000s, and the US reserves in the 1970s. China does not possess, as the United States does, vast deposits of surface-minable coal. More than 90% of China's coal comes from underground mines that can be as much as 1,000 metres deep, presenting increasing engineering challenges. We strongly suspect that the current reserves figures are too optimistic. The coal is certainly there, but — like the majority of coal elsewhere in the world — most of it is probably destined to stay put.

One way to estimate future production is to look at past production trends. This method was pioneered by geophysicist King Hubbert, who used 1950s data from the US oil industry to predict that US oil production would peak in the early 1970s. It did. Hubbert production profiles plotted over time assume the shape of a distorted bell curve, with a short peak and gradual decline (see graphic). Applying

Hubbert analysis to coal, Chinese academics Tao and Li<sup>7</sup> forecast in 2007 that China's production will peak and begin to decline long before the simple 62-years estimate, perhaps as early as 2025. During and after the period when production peaks, resource quality will dwindle and mining costs will rise, pushing up coal prices, as is already beginning to happen with Asia-Pacific coal.

Tao and Li used the Chinese government's latest official reserves figure of 187 billion tonnes to arrive at their peaking date between 2025 and 2032. Other forecasts are more pessimistic. A 2007 forecast<sup>3</sup> by the Energy Watch Group, based in Berlin, used a reserves figure of 114.5 billion tonnes (reported by China to

*"China's influence on future coal prices should not be underestimated."*

World Energy Council in 1992) to forecast a peak of production in 2015, with a rapid production decline commencing in 2020.

Analogous concerns raised in 1998 about the end of cheap oil<sup>8</sup> proved prescient. The price of oil has grown substantially since then, as have the costs of finding and extracting new supplies. The current price of more than US\$80 per barrel is about three times higher than the upper range in official forecasts for 2010 that were being issued in the late 1990s<sup>9</sup>. New technologies have made marginal oil reserves accessible, but deepwater drilling and oil-sands production entail high costs and risks.

Similarly, new technology — underground coal gasification — may eventually make marginal coal reserves accessible, but it will take time and substantial investment to commercialize on a large scale. Meanwhile, the world's highest-quality and most-accessible coal reserves are disappearing as demand for the fuel grows.

Coal consumption is accelerating fast, notably in China (see graphic). This renders meaningless reserves-lifetime figures calculated on the basis of flat demand. A 2009 report from China's Energy Research

Institute forecast that coal demand would rise by 700 million to 1 billion tonnes by 2020, reducing the reserves lifetime to about 33 years. If coal demand grows in step with projected Chinese economic growth, the reserves lifetime would drop to just 19 years<sup>10</sup>.

### COAL RELIANT

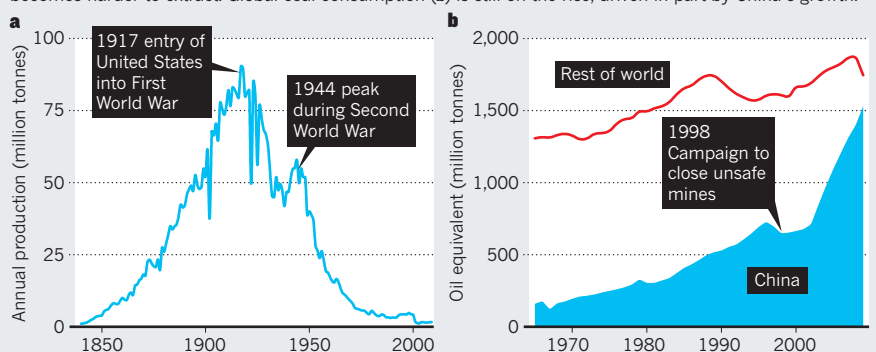
China has few options for reducing its reliance on coal. It uses coal in many more industries than the United States, where coal mostly fuels power generation. About half of China's coal provides 80% of the country's electricity supply; another 16% supplies the coke for its iron and steel industry, the largest in the world. Hundreds of millions of people in northern China consume another 6% for their winter heat supply. The remaining 28% is primarily used in industries such as cement, non-ferrous metals, and chemicals. Although China is rapidly expanding its supply of natural gas, to replace just the coal used for heating would double its total gas consumption.

Urbanization is also driving demand for coal. Less than half of China's population now lives in cities (compared with 80% for the United States and the European Union). To improve living conditions and opportunities for its citizens, the government wants the urban population to grow by 350 million people over the next 15 years, all of whom will require infrastructure such as housing, energy, transport, water supply and waste treatment. This will necessitate a steady supply of building materials such as cement, steel, aluminium and copper, all of which depend on coal for their production. Over the next decade, economic growth and urbanization are expected to use at the very least 700 million tonnes of coal — assuming that aggressive energy-efficiency and alternative-energy targets are also met<sup>7</sup>.

Can China go elsewhere for its coal? The United States has the world's biggest reported reserves, but almost all its current production — 1 billion tonnes — is used

### PRODUCTION AND CONSUMPTION

The annual production of coal in Pennsylvania (a) has been falling since the First World War as coal becomes harder to extract. Global coal consumption (b) is still on the rise, driven in part by China's growth.

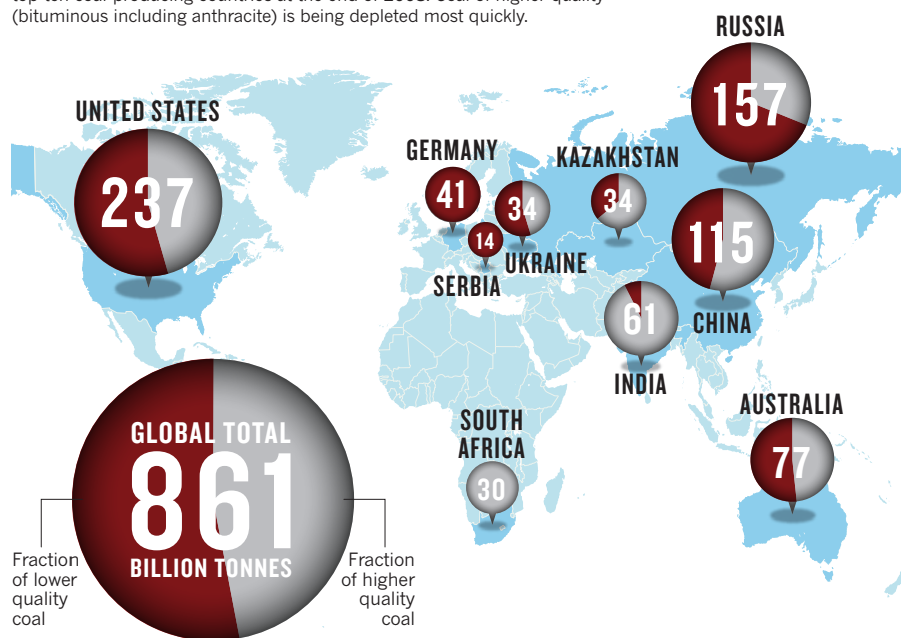


SOURCES: USGS, EIA, D. RUTLEDGE

SOURCE: REF. 6

# WORLD COAL RESERVES

Proven recoverable coal reserves reported to the World Energy Council by the top-ten coal-producing countries at the end of 2008. Coal of higher quality (bituminous including anthracite) is being depleted most quickly.



domestically. The biggest exporters of coal, Australia, Indonesia and South Africa, have much smaller reserves and production rates — some 250 million to 400 million tonnes a year. In 2008 the entire seaborne trade in steam coal (mainly used by power plants) amounted to about 630 million tonnes. Although this could grow (Australia, Russia and Indonesia are expanding capacity), growth will be limited, and prices pushed up, by the need to construct mines, railways and ports.

Russia has large but mostly undeveloped coal resources in Siberia. They are not located near demand centres, and rail transport of coal is expensive (which is why the largest exporters are coastal and trade is waterborne). Nevertheless, Russia could export Siberian coal to China more easily than to Europe, especially if China helped to build the railways.

China alone could absorb all current Asia-Pacific exports with just three years of import growth at current rates. Because other countries in the region also depend on coal imports, China clearly cannot take all, but competition for imports drives up prices. And then there's India, where imports are expected to nearly double to 100 million tonnes by 2012. India is one of the few countries to revise its reserves estimates upwards in recent years, but its higher-quality reserves are limited and it is importing increasing quantities.

The inevitable result of soaring demand and dwindling supply will be rising coal prices globally, even in

nations that are currently self-sufficient in the resource.

The poor quality of coal data globally means that uncertainty clouds every forecast. Even in the technologically advanced United States — the 'Saudi Arabia of coal' — most experts rely on decades-old coal surveys. These are commonly interpreted as indicating that the nation has a coal supply with a 250-year lifetime. This figure is not reliable enough for strategic energy planning.

In terms of energy output, US coal production peaked in the late 1990s (volume continued to increase, but the coal was of lower energy content). In 1995 the US Geological Survey (USGS) promised a new national coal survey, but it has not been seen as a high priority by that organization or by Congress. The most recent surveys<sup>11,12</sup> of two key mining regions show rapid depletion of high-quality reserves. There is still an enormous amount of US coal, but whether future energy production can be increased is doubtful, even taking into account new mining areas in Montana, Alaska and the Illinois basin.

## LIMIT CONSUMPTION

At the very least, the USGS should urgently complete a new national coal survey. And it is essential for the security of energy supplies globally that Chinese domestic coal production and the timing of its likely decline is better understood.

We believe that it is unlikely that world energy supplies can continue to meet projected demand beyond 2020. Therefore, new limits on energy consumption will be essential in all sectors of society — including

agriculture, transportation and manufacturing — and will be imposed by energy prices and shortages if they are not achieved through planning and policy.

Supply limits also have implications for the development of clean-coal technology. Also known as carbon capture and storage (CCS), clean coal is one proposal for reducing greenhouse-gas emissions while growing energy supplies. Because maintaining economic growth while cutting coal out of the energy equation globally will be difficult, and because nearly everyone assumes that coal will remain cheap far into the foreseeable future, the idea is to keep the carbon dioxide produced by burning coal from going into the atmosphere.

There are two hitches: the difficulty of scaling up such an enterprise, and its effect on electricity prices. As many analysts have noted, the scale and cost of clean-coal infrastructure will be vast<sup>13</sup>. Energy analysts agree that this will boost the price of electricity, but the scheme could work if coal prices remain low. If they don't, building new coal plants — conventional or clean — makes little economic sense, except to replace ageing inefficient infrastructure.

Nations should immediately begin to plan for higher fossil-fuel prices and to make maximum possible investments in energy efficiency and renewable-energy infrastructure. Even then the world will have to accept a slowdown in economic growth. ■

**Richard Heinberg and David Fridley** are at the Post-Carbon Institute in Santa Rosa, California 95404, USA. Heinberg is the author of nine books, including *Blackout: Coal, Climate, and the Last Energy Crisis*. e-mail: richardheinberg@postcarbon.org

1. Patzek, T. W. & Croft, G. D. *Energy* **35**, 3109–3122 (2010).
2. Mohr, S. H. & Evans, G. M. *Fuel* **88**, 2059–2067 (2009).
3. Zittel, W. & Schindler, J. Energy Watch Group, Paper No. 1/07 (2007); available at <http://go.nature.com/jngfsa>
4. Rutledge, D. *Hubbert's Peak, The Coal Question, and Climate Change* (2007); available at <http://rutledge.caltech.edu>
5. Höök, M., Zittel, W., Schindler, J. & Aleklett, K. *Fuel* **89**, 3546–3558 (2010).
6. 2010 Survey of Energy Resources (World Energy Council, 2010); available at <http://go.nature.com/hde5r7>
7. Tao, Z. & Li, M. *Energy Pol.* **35**, 3145–3154 (2007).
8. Campbell, C. J. & Laherrère, J. H. The End of Cheap Oil. *Sci. Am.* (March 1998).
9. Energy Information Administration. *Annual Energy Outlook 1998* (DOE/EIA, 1997).
10. 2050 China Energy and CO<sub>2</sub> Emissions Report (in Chinese) (Science Press, 2009).
11. Luppens, J. A. et al. *Assessment of Coal Geology, Resources, and Reserves in the Gillette Coalfield, Powder River Basin, Wyoming*. Open-File Report 2008-1202 (USGS, 2008).
12. *Coal Reserves of the Matewan Quadrangle, Kentucky — A Coal Recoverability Study*. US Bureau of Mines Circular 9355 (USGS, 2003).
13. *Strategic Analysis of the Global Status of Carbon Capture and Storage*. (Global CCS institute, 2009).

➔ **NATURE.COM**  
For more on low-carbon energy see:  
[go.nature.com/ui4npe](http://go.nature.com/ui4npe)





# Questioning economic growth

Our global economy must operate within planetary limits to promote stability, resilience and wellbeing, not rising GDP, argues **Peter Victor**.

The idea that governments of developed countries should no longer pursue economic growth as a primary policy objective is widely regarded as heresy. Yet a growing number of scholars, policy-makers and citizens are coming round to the idea that the planet cannot sustain continued global economic growth. Even economist Robert Solow, who won the 1987 Nobel Prize in Economics for his work on economic growth, said in 2008 that the United States and Europe might soon find that “either continued growth will be too destructive to the environment and they are too dependent on scarce natural resources, or that they would rather use increasing productivity in the form of leisure”<sup>1</sup>. The idea of steady-state economies, or even economic ‘degrowth’, in developed countries is gaining traction.

The reasons for disenchantment with

economic growth as a paramount policy objective are not hard to find. Humanity has gone beyond the ‘safe operating space’ of the planet with respect to climate change, nitrogen loadings and biodiversity loss, and threatens to do so with six other major global environmental issues<sup>2</sup>. This excessive burden on Earth can be traced to the massive increase in the materials, fossil fuels and biomass used by the world’s economies. Mankind’s ‘throughput’ — the sheer weight of materials, including fuel, that feed the world’s economies — has increased 800% in the twentieth century<sup>3</sup>, with a correspondingly large increase in wastes returned to the environment. In the same time, the human population has risen from 1.6 billion to more than 6 billion, and our presence has been felt over an increasingly large part of Earth’s surface. All of this drove and was driven by

unprecedented economic growth, the benefits and costs of which have been spread remarkably unevenly around the planet.

A key question now is whether and how economies can develop in a way that respects Earth’s biophysical boundaries and feeds the 9 billion people expected by mid-century.

One option is for developed countries to continue striving for economic growth, while attempting to reduce impacts on the planet. This means betting that economic growth can be successfully and rapidly decoupled from material and energy inputs. Such ‘green growth’ is currently favoured by the Organisation for Economic Co-operation and Development (OECD). But it can be confounded by the rebound effect: efficiency improvements often induce changes that reduce, nullify or outweigh environmental and resource benefits. This was first recognized in 1865 by economist W. S. Jevons, who noted that improvements in steam engines were accompanied by an increase in total coal consumption.

By 1910, the best steam engines in the United Kingdom were about 36 times more efficient than those of 1760 (ref.4), but a 2,000-fold rise in steam-power use<sup>5</sup> had increased coal consumption dramatically. A rebound of 50% is not unusual for many technologies.

## WHAT PRICE HAPPINESS?

An alternative is to encourage growth in sectors of the economy that use fewer resources, such as the service sector. Such a strategy could buy some time, but not if it simply shifts the production of resource-intensive products and their related environmental burdens to other countries, as has been the pattern in recent years.

A third option is to limit growth itself. The battle against climate change illustrates the attractiveness of this strategy. To reduce greenhouse-gas emissions (GHG) by 80% over 50 years, an economy that increases its real gross domestic product (GDP) by 3% a year must reduce its emissions intensity — tonnes of GHG per unit of GDP — by an astonishing 6% a year. For an economy that does not grow, the annual cut would be a still very challenging 3.2%.

The view that we should curb planetary impacts by reducing growth in richer countries is reinforced by several considerations. First, there is mounting evidence that this growth is largely unrelated to measures of happiness. Second, in recent decades, increasing inequality has accompanied much of this growth, leading to problems ranging from poor public health to social unrest. Third, the prospects for real improvement in the developing world are likely to be diminished if developed countries continue to encroach on more ecological space.

Removing economic growth as a major

policy priority runs counter to the views of governments and many international agencies. Many nations responded to the recent financial crisis with desperate measures to resume economic growth. Yet when we recognize how briefly economic growth has held such prominence in policy circles, dethroning it seems less improbable. Regular estimates of GDP by governments date back only to the 1940s, and the measure was initially used in support of specific objectives, such as stimulating employment. Only in the 1950s did economic growth become a policy priority in its own right<sup>6</sup>.

Economists and other social scientists now need to map out functional economies in which growth is sidelined, and stability, resilience and wellbeing are the prime objectives, within environmental and resource constraints. Ecological economist Herman Daly, who has investigated and promoted a steady-state economic model for several decades, has formulated a useful set of principles for limiting material use, including: the harvest of renewable resources should not exceed their regeneration rate; the rate of extraction of non-renewable resources should not exceed the rate of creation of renewable substitutes; and waste emissions should not exceed the environment's capacity to assimilate them. To these we should add the protection of land and water to reduce competition among humans and other species. Among the many successful applications of these principles is the creation of protected areas and green belts.

Daly, with theologian John Cobb, also proposed an alternative measure of macro-economic success: the Index of Sustainable Economic Welfare (ISEW), incorporating environmental degradation, resource depletion and other factors. Estimates of this index show a major divergence from GDP per person for many countries. In one study by environmental charity Friends of the Earth<sup>7</sup>, the gap between US GDP and the 'Genuine Progress Indicator' (GPI), calculated similarly to the ISEW, was particularly marked: whereas GDP per person rose from the 1970s, GPI actually declined (see 'Genuine progress?').

### SHORTER WORK YEAR

These results bear out an observation made in 1934 by Simon Kuznets, a Russian-American economist and one architect of the system of national accounts from which GDP is derived<sup>8</sup>: "The welfare of a nation can scarcely be inferred from a measure of national income." Work on more broad-based indicators to complement or replace GDP has been given a substantial boost by a 2009 report by Nobel laureates Joseph Stiglitz and Amartya Sen<sup>9</sup> that caught the attention of many politicians.

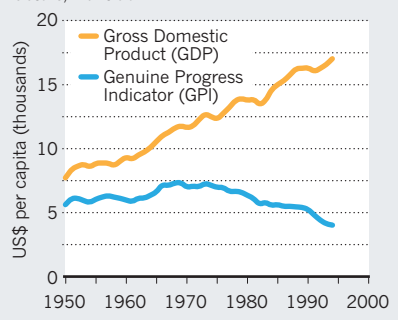
Models have been built to explore what

might realistically be accomplished in developed countries that forgo economic growth, and what the consequences might be. I constructed<sup>10</sup> a fairly conventional model of the Canadian economy and found circumstances under which employment can be increased, poverty and greenhouse-gas emissions reduced, and government debt effectively managed without economic growth. A key ingredient is a shorter work year, which would help to spread employment among more of the labour force. The benefits of greater productivity would thus be directed towards more leisure time, rather than increasing GDP. Scoping this out for Canada, assuming that labour productivity continues to rise modestly, a reduction in the average work year of around 15% by 2035, to 1,500 hours a year, would secure full employment. This work year would still be longer than in some European countries. In Germany, for example, the average paid employee worked 1,430 hours in 2008.

Other ingredients for an attractive low/no-growth scenario include more focused and

### GENUINE PROGRESS?

US GDP rose over the past decades; the GPI, which accounts for social and environmental factors, went down.



better-funded anti-poverty programmes, a stable population (already achieved in many developed countries and within the grasp of others), and stricter policies on environment and resources, based on Daly's principles. My study has helped to stimulate similar investigations, under way or proposed, in countries including New Zealand, Austria, the United Kingdom, Finland and the United States, with results expected over the next year or so.

Zero economic growth, however, may not be enough. Some researchers are looking seriously at 'degrowth': shrinking developed economies to bring them into balance with resource and environmental limits, while improving quality of life. The scope of changes in all aspects of the economy would be much more far-reaching, and the repercussions for society greater. Nevertheless, degrowth in materials use, fossil energy, land and water is clearly required, so degrowth of national economies may be unavoidable.

There is debate about whether capitalism

is compatible with steady-state or degrowth economies. A shrinking economy brings a real risk that profit-seeking companies and their shareholders will be disappointed, credit ratings will suffer, the financial system will be in jeopardy, trade will shrink and the whole capitalist system could spiral to collapse. Whether this would happen remains an open question. Solow, for one, sees no reason why capitalism could not survive with slow or even no growth. Others are more sceptical — especially about the survival of capitalism in degrowth societies. It is worth noting that even in a shrinking economy, some sectors — such as renewable-energy development — will flourish.

As long as economic growth remains so important to global policymakers, humanity is hopelessly constrained: the environmental policies we need face the unreasonable political hurdle that they must also be shown to promote economic growth. This must change. At grass-roots level, many people in the developed world are already directing their energies towards enhanced wellbeing, in part by turning to local producers for their food, clothing and other needs. Institutions of all kinds — financial, political, legal, educational, religious and social — that have evolved to thrive in a fast-growing economy will have to adapt. This could be the greatest challenge of all; there are no good answers yet as to how they should change.

With the prospect of environmental calamity facing humanity, developed economies must chart a course towards living within a fair share, and no more, of the planet's safe operating space. Developing countries, in their turn and time, will also have to adjust. Done thoughtfully, this could lead to more satisfactory and fulfilling lives for all. ■

**Peter Victor** is an economist at York University in Toronto, Ontario and author of *Managing Without Growth: Slower by Design, Not Disaster*.  
e-mail: [pvictor@yorku.ca](mailto:pvictor@yorku.ca)

1. Stoll, S. *Harper's Magazine*. <http://www.harpers.org/archive/2008/03/0081958> (March 2008).
2. Rockström, J. et al. *Nature* **461**, 472–475 (2009).
3. Krausmann, F. et al. *Ecol. Econ.* **68**, 2696–2705 (2009).
4. Smil, V. *Energy in World History* (Westview Press, 1994).
5. Crafts, N. *Econ. J.* **114**, 338–351 (2004).
6. Arndt, H. W. *The Rise and Fall of Economic Growth* (Longman, 1978).
7. Friends of the Earth. Indexes of Sustainable Economic Welfare. <http://www.foe.co.uk/community/tools/isew/international.html>.
8. Kuznets, S. *National Income, 1929–1932*. Senate Document 124. <http://library.bea.gov/u/?SOD,888> (US Congress, 1934).
9. Stiglitz, J., Sen, A. & Fitoussi, J.-P. *Report by the Commission on the Measurement of Economic Performance and Social Progress* (French Commission on the Measurement of Economic Performance and Social Progress, 2009).
10. Victor, P. A. *Managing without Growth: Slower by Design, Not Disaster* (Edward Elgar, 2008).





Technologies such as this drone are becoming increasingly independent of humans.

## SOCIETY

# The rise of the 'technium'

Kevin Kelly argues compellingly that technology is taking on a life of its own, finds **Zaheer Baber**.

In *What Technology Wants*, writer Kevin Kelly radically rethinks the relationship between humans and technology. Scientific inventions have become so complex and interwoven with our lives, he says, that humans have less and less sway over how mechanical systems evolve. Nor can we stop the spread of technologies. Consequently, when assessing future risks, we should adopt a proactive approach of trial and error and revision, rather than strict precaution.

To make his point, Kelly introduces the

concept of the 'technium' to embody the vast techno-social system. Distinct from individual innovations such as radar or plastic polymer, the technium includes all the machines, processes, society, culture and philosophies associated with technologies. The sheer complexity of interactions between the various layers and loops of the technium gives it a degree of autonomy. As it evolves, it develops its own dynamics.

According to Kelly, an autonomous system displays traits of self-repair, self-defence,

self-maintenance, self-control and self-improvement. No current system has all these properties, he admits, but many technologies exhibit some of them. Aeroplane drones can self-steer and stay aloft for hours, for instance, but cannot repair themselves. Communication networks can repair themselves but cannot self-reproduce. Computer viruses can self-reproduce but cannot improve themselves. As technologies multiply and become more adaptive, the technium is becoming increasingly autonomous.

For example, the vast global communications network incorporates 170 quadrillion computer chips (a quadrillion is  $10^{15}$ ) wired up into one giant computing platform, with a density of links approaching that of synapses in the human brain. Scientists can trace the majority of traffic flowing through the networks, but occasional bits are lost or transformed during transmission. Most of these mutations of information are attributable to causes such as hacking and machine error, but a few per cent are not — these changes originate not from humans, but from vagaries in the system itself. The flow of bits through the telephone network has, in the past decade, become statistically similar to the fractal pattern found in self-organized systems. This suggests that it is developing behaviour of its own.



**What Technology Wants**

KEVIN KELLY  
Viking: 2010. 416 pp.  
\$27.95

Although the technium has neither an idea of self nor conscious desires, it develops mechanical tendencies, or 'wants', through its complex behaviour. Its millions of amplifying relationships and circuits of influence push the technium in certain directions. For example, some personal robots can navigate obstacles to seek out power outlets and plug themselves in to be recharged. For Kelly, these robots are like bacteria drifting towards nutrients with no conscious awareness of that goal. As frontier technologies increase in sophistication, these 'wants' gain in both complexity and force. Moreover, the tendencies become increasingly independent of human designers and users.

As Kelly points out, technophobes and technophiles alike agree that the technium is spinning beyond human control. They disagree only on what should be done about it: whether the technium should be stopped, modified or embraced. Kelly respects all

**NATURE.COM**

For more on  
the evolution of  
technology, see:  
[go.nature.com/tysg2k](http://go.nature.com/tysg2k)

sides in the polarized debates about technology. He accepts the unease that the technium can unleash, devoting chapters



to the anti-technology manifesto of the Unabomber, Ted Kaczynski, and the selective uptake of innovations by Amish people. He recognizes those who have positions in between, including the proponents of indigenous knowledge and inventors themselves.

Apprehension about the technium assuming a life of its own continues to grow with the rise of genomics, robotics, informatics and nanotechnology. Cautious states and

**“Every technology produces degrees of good, harm and risk, and the evolution of each is uncertain.”**

publics often turn to the precautionary principle, which holds that any technology must be shown to do no harm before it can be embraced. Kelly argues that this approach is impractical, unfeasible and unattainable. Every technology produces

degrees of good, harm and risk, and the evolution of each is uncertain — none can ever be said to be decisively safe.

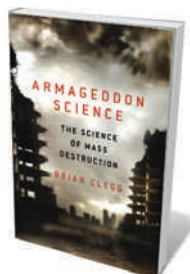
As an alternative, Kelly draws on philosopher Max More's 'proactionary principle', which states that the only way to evaluate new technologies is to try them out as prototypes and then refine them. To evaluate risk we must continually assess new technologies in the context of use. Kelly pares More's principle down to five elements: anticipation; continual assessment; prioritization of risks; rapid correction of harm; and redirection.

Owing to the autonomy of the technium, Kelly contends, it is pointless to ban risky technologies. Attempts to put a moratorium on them will only ensure that the emergent ones will be even more impervious to human control — exhibiting a form of natural selection. Instead, we should strive to produce technology that is 'more convivial' — that is, more compatible with life. Kelly believes that every technology can be channelled towards uses that promote greater transparency and more collaboration, flexibility and openness across society.

He draws extensively on other studies, particularly Langdon Winner's groundbreaking book *Autonomous Technology* (MIT Press, 1977). Winner famously discussed uncontrollable “technological drift” as one of the most disturbing features of modern life. He also extensively used the phrase ‘socio-technical system’ rather than ‘social system’ to capture the seamless amalgamation of humans and technology. But Kelly's concept of the technium and his description of how it attains autonomy are original and timely. ■

**Zaheer Baber** is professor of sociology at the University of Toronto, Ontario, M5S 1A1, Canada, and author of *CyberAsia*. e-mail: zaheer.baber@gmail.com

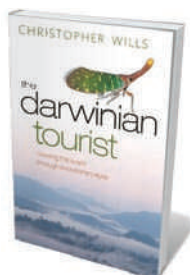
## Books in brief



### **Armageddon Science: The Science of Mass Destruction**

Brian Clegg ST MARTIN'S PRESS 304 pp. \$25.99 (2010)

From biohazards to climate change, there are many ways to erase humanity. Physicist Brian Clegg assesses a range of doomsday scenarios in his book. Although he remains unshaken by the rumoured risk of miniature black holes being created by the Large Hadron Collider at CERN, Europe's particle-physics lab in Geneva, Switzerland, he accepts that nanobots and nuclear technologies are credible threats. Ultimately, he is an optimist, who hopes that better science education will help us to make the best choices about our future.



### **The Darwinian Tourist: Viewing the World Through Evolutionary Eyes**

Christopher Wills OXFORD UNIVERSITY PRESS 288 pp. \$34.95 (2010)

In this travel book with an evolutionary bent, biologist Christopher Wills relates his personal journeys to the world's wildest places. He describes the biodiversity of the Peruvian rainforest, and meets wolf cubs in a Mongolian village to reveal how the domestication of dogs began. He goes on to piece together the story of human evolution with that of the hunter-gatherer peoples of the African Kalahari and the bones of ancient hominins in the island caves of Flores, Indonesia.



### **Living with Complexity**

Donald A. Norman THE MIT PRESS 280 pp. \$24.95 (2010)

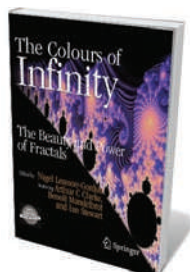
The complexity of modern technology is a boon rather than a problem, argues influential designer Donald Norman. Just as the owner of a messy desk can quickly locate papers in seemingly random piles, even the most difficult technologies can be tamed through good design and mastery. He sees this as a partnership between the designers who produce objects that tame complexity, and the consumers who must learn the skills needed to use those innovations. Once under control, the cleverest technologies may become as easy to use as a pencil or salt shaker.



### **Gregory Petsko in *Genome Biology*: The First 10 Years**

Gregory Petsko BIOMED CENTRAL 304 pp when printed. Available for Kindle (\$1.13) and on iPad and iPhone (free) (2010)

To mark the tenth anniversary of the launch of the journal *Genome Biology*, publisher BioMed Central is releasing an e-book compilation of the columns of structural biologist Gregory Petsko, who has written for the journal every month since 2000. With his characteristic wit and perception, Petsko muses on trends in genomics research, funding and policy. He also discusses how the culture of genomics research is changing in an era of blogs and social networks.



### **The Colours of Infinity: The Beauty and Power of Fractals (Second Edition)**

Edited by Nigel Lesmoir-Gordon SPRINGER 174 pp. \$59.95 (2010)

In 1995, a groundbreaking television documentary introduced the world to the psychedelic geometries of the Mandelbrot set. Essays on the beauty and mathematics of fractals by the film's contributors, including the father of the field, Benoît Mandelbrot, are collected in Nigel Lesmoir-Gordon's book. This updated edition includes a new chapter written by Mandelbrot just before his death, and the 1995 documentary has been remastered on an associated DVD.

# Greenhouses in the sky

Emma Marris is intrigued by an optimistic vision of high-rise farms.

To feed the cities of the future, ecologist and microbiologist Dickson Despommier envisions a global shift to indoor agriculture. In *The Vertical Farm*, he sets out his big idea: by raising crops and animals in large urban buildings, former farmland can be returned to forest and rivers can be spared poisonous run-off. Cities will no longer have to transport food in and waste out, but will be self-sustaining — the urban equivalent of a natural ecosystem.

Despommier, who has developed and promoted his concept for a decade, imagines filling skyscrapers with hydroponically or aeroponically grown crops, medicinal herbs and biofuels fed on “ultrapure, chemically defined diets”. Zones of plants would be dedicated to filtering the city’s waste water back into drinkability. Every neighbourhood, rich or poor, would have access to wholesome, tasty, local food.

His idea is enthralling — but far from

realized. It brings to mind the urban plans of King Camp Gillette, inventor of the safety razor, who sketched out a high-rise utopia in his 1894 book *The Human Drift*. His Metropolis was to be an enormous white porcelain city of “immaculate cleanliness”, powered by Niagara Falls and run by machines, that would house the entire population of the United States. Gillette didn’t bring agriculture within its walls — his vision was to import raw crops from surrounding lands and process them centrally. Despommier makes no mention of food processing in *The Vertical Farm*; perhaps in the future everyone cooks from scratch. He’s clearly an optimist.

Like Gillette, Despommier has society’s best interests at heart, but his proposal is grandiose. A professor of public health, he too is obsessed with cleanliness and offers discourses on the health threats of human faeces, rats and other vermin and the various contaminants that vertical-farm workers will



**The Vertical Farm: Feeding the World in the 21st Century**

DICKSON

DESPOMMIER

Thomas Dunne Books:  
2010. 320 pp. \$25.99

take steps to block. He makes easy assumptions; for example, that labourers will submit to regular blood tests to prevent the spread of disease, wear sterile uniforms and live on the premises. And that “in most cases”, abandoned farmland around the world will become hardwood forest. Most outlandish is his blazing confidence in his idea. In ten

years of study, he tells us, he has thought of “no significant disadvantages” to his scheme, except the minor matters of construction costs and farmer displacement.

One downside is easy to spot: the massive amounts of energy required to grow plants indoors. Police often bust major marijuana-growing operations by following up on unusually high electricity bills. Using coal or gas to grow strawberries and tomatoes is a lot more expensive than energy from the Sun. Despommier addresses this with transparent buildings and technology: light-emitting diodes on flexible plastics wrapped around individual plants, mirrors, solar panels, wind turbines and plasma-arc gasification facilities to turn biological waste into energy.

Yet his detail, he admits, doesn’t extend to a quantitative demonstration that vertical farms are competitive in terms of either energy or money. He saves both by eliminating conventional farm machinery, pesticides, herbicides, fertilizer, transport and other costs, including crop failure. But he spends energy and money by building tall, complex buildings in urban cores and in acquiring and maintaining cutting-edge infrastructure, airtight security and negative-pressure ventilation.

Maybe growing all of our calories within the city limits is no more likely than thinking we’ll all move to Metropolis. But *The Vertical Farm* is nevertheless inspiring. For some crops in some places, it might make sense. If Despommier won’t do the maths, someone should. Any idea that might help us to avoid displacing any more natural areas with agriculture deserves a hearing. ■

Emma Marris writes for *Nature* from Columbia, Missouri.



Bringing farms into city buildings might save on transport, but energy costs could skyrocket.





Kristen Bush as Rosalind Franklin in a play about collaboration, competition and pursuing scientific glory.

## HISTORY

# Franklin, centre stage

Josie Glausiusz enjoys a play capturing the zeal and backstabbing in the race to discover DNA's structure.

A bell chimes, and a moment of calm descends on the stage, as Rosalind Franklin marvels at an X-ray diffraction image. "It's a perfect X. It's a helix," she says. "I've never seen anything like it."

What Franklin saw that night in May 1952 is at the heart of Anna Ziegler's powerful new play, *Photograph 51*, funded by the Alfred P. Sloan Foundation and now showing at the Ensemble Studio Theater in Manhattan. Played in plummy tones by Kristen Bush, Franklin is the focus of this fast-paced performance, which dramatizes the obsessive and, at times, devious race to discover the structure of DNA.

Ziegler was originally commissioned by a Maryland theatre, Active Cultures, to create a play about three women scientists, but rewrote the script to focus on Franklin alone after realizing that it was her story that really grabbed her. As a Jewish woman, Franklin was thwarted by obstacles of the time — sexism and anti-Semitism — and by her own internal limitations, Ziegler says. Her toughness got her where she was, but it also meant that she guarded her ideas from outside interference. "The play is largely about Franklin's inability to collaborate, or lack of

## Photograph 51

WRITTEN BY ANNA ZIEGLER; DIRECTED BY LINSAY FIRMAN.

*The Ensemble Studio Theatre, New York City. Until 21 November 2010.*

crystallized molecules of DNA. It was later shown without her knowledge to James Watson, who recognized the helix as the missing piece of the puzzle that enabled him and his collaborator Francis Crick to construct their famous model of the molecule of life. With a cast of characters that includes a wild-haired Watson (played by Haskell King) and a bewildered Maurice Wilkins (Kevin Collins) — Franklin's colleague at King's College London, who revealed the image to Watson — Ziegler has produced a witty and poignant account of the controversy surrounding DNA's discovery.

The play is based on fact, and in large part on Brenda Maddox's moving biography, *Rosalind Franklin: The Dark Lady of DNA* (HarperCollins, 2002), as well as Watson's best-selling *The Double Helix* (Atheneum, 1968). Ziegler crams a great

desire to," she adds.

The famous Photograph 51 from which the play takes its name is Franklin's best recording of the patterns produced by bouncing X-rays off

deal of complicated science into 90 minutes, capturing both the zeal and the backstabbing that often accompany the pursuit of scientific glory. The compressed format, however, precludes the wealth of detail that appeared in Maddox's biography, which presents Franklin not just as a dedicated scientist but as an elegant and generous young woman with many friends, who loved hiking and travelling.

As in real life, Franklin battles in the play for respect as an accomplished woman scientist in the 1950s. She faces a barrage of belittling rules and remarks, not least from Watson who, as he wrote in *The Double Helix*, wondered how "Rosy" would look "if she took off her glasses and did something novel with her hair." The characters in the play inform Franklin, like a chorus of bad fairies, that she might have achieved more if she had been more open, less wary, willing to take more risks, make models, move forwards without the certainty of proof. She may have triumphed, they jibe, if she was born at another time — or born a man.

The sexism that Franklin faced in the stifling environment of King's may explain why the playwright depicts her as stubborn, secretive and rarely happy. One of her few moments of serenity comes during an imagined conversation with a close friend, US biophysicist Don Caspar. Asked what she wants, she replies, "So many things: to wake up without feeling the weight of the day pressing down ... to eat more beets and also turnips, to be kissed ... be a child again, held up and admired, the world full of endless future."

Alas, Franklin's future was cut short by ovarian cancer, from which she died in 1958. Although not covered by the play, she had moved in 1953 to Birkbeck College, now part of the University of London, where she worked happily on the structure of tobacco mosaic virus. Because the Nobel prize is not awarded posthumously, she did not share in the 1962 prize in physiology or medicine that was awarded to Watson, Crick and Wilkins. But as Ziegler conveys, it is not clear that she was fixated on the prize, although no doubt she would have been happy to win it.

"She was more about the work and the process, and not as much about the accolades," Ziegler says. "It was about the personal satisfaction of understanding and cracking something. She was in a different kind of race." ■

"She may have triumphed, they jibe, if she was born at another time — or born a man."

Josie Glausiusz is a journalist based in New York and a contributor to the science blog [www.lastwordonnothing.com](http://www.lastwordonnothing.com).





In *Monkeys as Judges of Art* (1889), Gabriel von Max conveys his interest in animal and human nature.

## PAINTING

# Inquisitive and exact

**Alison Abbott** visits an exhibition charting the artistic and scientific interests of painter and collector Gabriel von Max.

During his lifetime, Gabriel von Max (1840–1915) was one of Munich's most successful artists — a privilege he exploited. In his middle-age, he began to churn out reams of paintings for the art market, for he had a very expensive habit to feed: collecting scientific objects.

His commercial tendency and his adherence to a realistic painting style might explain why he fell into obscurity as art moved on in the twentieth century. An exhibition now on at the Kunstbau gallery in Munich, Germany, claims to rediscover this extraordinary man, who studied and painted nature with the inquisitiveness and exactness of Leonardo da Vinci while embracing the radical new sciences of his age with equal passion.

The show brings together both sides of his psyche: the artistic and the scientific. It displays a broad range of his paintings — from early religious works to later studies of primates and commentaries on the scientific process — alongside objects from his collection. At his death, his acquisitions totalled up to 80,000 objects, including around 400 skulls believed to have been destroyed in the Second World War, but which were rediscovered in Freiburg, Germany, in 2008.

His earlier paintings were concerned with religion or death, and conveyed a heightened emotionality along with a teasing eroticism.

**Gabriel von Max: Star Artist, Darwinist, Spiritualist**  
Kunstbau, Munich.  
Until 30 January 2011.

His breakthrough came with his 1867 work, *Christian Martyr on the Cross* (St Julia), a luminous painting of such power that female visitors to its first showing openly wept, according to reports at the time. Even then, his fascination with nature was on display. He painted a fly or butterfly motif into many pictures, settling them casually on a death-white arm or anatomy table. Those insects were far from casually painted, however, as his detailed preparatory sketches show.

Even more fascinating are his sketches and paintings of the monkeys he collected and kept as pets. A capuchin monkey called Paly was his constant companion for 15 years. His interest in — and affection for — the animals paralleled his embrace of Charles Darwin's theory of evolution. He even saw them as superior in some ways to humans, who he thought were corrupted by civilization.

At the turn of the century, von Max completed a series of paintings showing monkeys conducting academic activities such as giving anatomy lessons. The most familiar, *Monkeys as Judges of Art* (1889), which portrays 13 monkeys as art critics (pictured), is widely assumed to be a censure of the profession. Yet the artist's writings, according to the

exhibition's catalogue, suggest the opposite. He sought to convey sophisticated, individual human weaknesses — such as vanity — just as writers of fables traditionally used particular animals to embody human characteristics.

Von Max may have loved his pets, but he studied their behaviour and anatomy with detached scientific rigour. Many died in the cold Bavarian climate, and he would skin their bodies, sketching and photographing their muscles to understand how to portray postures correctly.

But living animals should not be harmed merely to satisfy scientific curiosity, cautioned von Max in his 1883 painting *The Vivisector*. The vivisector, a bearded scientist, sits at his dissection table. The allegorical female figure of compassion has taken from him a puppy, with bound muzzle, which he was preparing to dissect. The scales she holds aloft in her other hand show that the heart weighs more than the brain in this situation. That painting was quickly used as propaganda by the growing anti-vivisection movement, which was already putting Germany's physiologists and infection biologists on the defensive.

Von Max's scientific collection, replete with objects representing the new sciences of geology, ethnology, anthropology and palaeontology, reflected his life-long concern with the origins of humans and the Earth, and was taken seriously by the scientific community. Zoologist and artist Ernst Haeckel, who became a friend, engineered for him an honorary doctorate from his University of Jena in Germany.

After von Max's death, the collection was bought by the Mannheim museum. It was broken up in 1935 and distributed among specialist museums in the region — which is how the skulls ended up in the University of Freiburg's anthropology collection. While preparing for a German exhibition celebrating the 150th anniversary of Darwin's theory of evolution, curators found that von Max's skulls had not been destroyed in the Second World War after all, but had got mixed up with a different skull collection. Von Max's entire collection is now reassembled at Mannheim's Reiss-Engelhorn Museum.

As this fine exhibition shows, von Max is well worth bringing back into the light. ■

**Alison Abbott** is Nature's Senior European Correspondent.

## CORRECTION

In 'The light and shade of German science' (*Nature* **467**, 660; 2010), the date of the Berlin Wall's rise was incorrectly given as 1949, the date of Germany's separation into two states. Construction of the Berlin Wall began in 1961.

PHOTO SCALA, FLORENCE/BPK, BILDAGENTUR FUER KUNST, KULTUR UND GESCHICHTE, BERLIN

# CORRESPONDENCE

## European bounty for taxonomists

Non-professional taxonomists have been responsible for describing more than half of the animal species discovered in Europe from 1998 to 2007 (see also *Nature* 467, 788; 2010). The extraordinary current rate of description of new species makes Europe an unexpected frontier for biodiversity exploration.

The Fauna Europaea database ([www.fauna-eu.org](http://www.fauna-eu.org)), released in 2004, lists more than 125,000 European species of multicellular terrestrial and freshwater animals. More than 700 new species are described each year in Europe — four times the rate of two centuries ago. However, we have not yet reached saturation in the inventory of European fauna, and we cannot accurately estimate the total number of species living in the continent's ecosystems.

The unprecedented rate of species description has depended heavily on the scientific contribution of unpaid scientists (non-professional and retired professional taxonomists). More attention should be given to ways of enhancing this formidable workforce.

There is an urgent need for an effective policy-supported business plan to complete the biodiversity inventory at European and national levels, preferably targeting species-rich and less-charismatic groups such as mites, rove beetles, micro-wasps and nematodes. Amateurs could be readily integrated into such a framework of defined and coordinated objectives.

The future of amateur taxonomy also depends on incorporating molecular techniques, either through formal training or through collaboration between molecular-oriented

professionals and morphology-oriented citizen scientists.

**Benoît Fontaine** on behalf of 51 co-authors\*, *Muséum National d'Histoire Naturelle, France.*  
[fontaine@mnhn.fr](mailto:fontaine@mnhn.fr)

\*A full list of signatories is available online at <http://dx.doi.org/10.1038/468377a>

## Innovation in Europe — three questions

Three long-standing questions still need to be addressed to stimulate innovation in the European Union (*Nature* 467, 1005; 2010).

First, to what extent can governments make informed choices about which areas should be stimulated by public (and private) funding of research and development (R&D)? Governments generally lean towards areas with a strong past performance rather than favouring those with a promising future. Are public agencies — or any other organization — capable of picking future winners?

Second, assuming that governments have the capability and remit to select promising areas, the next question is whether the European Union is the proper level for policy interventions. To put it another way: to what extent do European-wide innovation partnerships yield better products than national or regional ones? This everlasting debate becomes even more relevant in the implementation and feasibility of large-scale R&D projects. Perhaps one should accept a variety of spaces for public R&D intervention — some sectors require international research and innovation policies, whereas others are the realm of regional policies.

Third, there is the issue of how to organize innovation

projects that address societal issues. You rightly point out the challenges of coordinating multiple-actor constellations. However, science and technology studies teach us that proactively involving stakeholders from different backgrounds and disciplines can be beneficial to the 'responsible' steering, utilization and implementation of R&D.

**Wouter Boon, Gaston Heimeriks** *Utrecht University, the Netherlands.*  
[w.boon@geo.uu.nl](mailto:w.boon@geo.uu.nl)

## Misreporting: a glowing report

As a former science writer for several UK national newspapers, I commend Simon Lewis for his balanced and valuable analysis of how to deal with misreporting (*Nature* 468, 7; 2010).

Lewis avoids the common error of assuming that the bylined journalist was responsible for the headline or the final text. As I know all too well, stories can be extensively rewritten without being referred back to the named author. Complaining about this practice is regarded as naive and career-limiting.

His experiences show how one can use the rivalries that exist between newspapers to obtain some redress for misreporting. Newspapers delight in reporting egregious examples of misreporting by rivals.

Thus, in approaching the UK newspaper *The Guardian*, Lewis targeted his complaint about the original *Sunday Times* report perfectly. I am glad that Lewis was able to gain some redress. I am also grateful to him for reminding me how good it is to be out of the newspaper business.

**Robert Matthews** *Aston University, UK.*  
[rajm@physics.org](mailto:rajm@physics.org)

## Reef technology to rescue Venice

Rachel Armstrong and Neil Spiller suggest that Venice's sinking foundations might be supported by an artificial reef grown using 'protocells' that precipitate limestone from sea water (*Nature* 467, 916–918; 2010). The technology already exists to grow structures rapidly from sea water, and this could be applied in Venice immediately.

'Biorock' electrolysis of sea water has been used for nearly 35 years in more than 20 countries to grow limestone structures of any size and shape in sea water and brackish water (W. Hilbertz *IEEE J. Oceanic Eng.* 4, 94–113; 1979).

Biorock products have a load-bearing strength of up to 80 newtons per square millimetre (80 megapascals), around three times higher than concrete made from ordinary Portland cement. Corals and oysters grow faster and survive environmental stress better on Biorock structures. These have helped to restore severely eroding beaches on atoll islands within just a few years (for example, see [go.nature.com/buyqjk](http://go.nature.com/buyqjk)).

**Thomas J. Goreau** *Global Coral Reef Alliance, Massachusetts, USA.*  
[goreau@bestweb.net](mailto:goreau@bestweb.net)

### CONTRIBUTIONS

Submissions to Correspondence may be sent to [correspondence@nature.com](mailto:correspondence@nature.com) after consulting the author guidelines at <http://go.nature.com/cmCHno>. They should be no longer than 350 words. Readers are also welcome to comment online on anything published in *Nature*: [www.nature.com/nature](http://www.nature.com/nature).



# Benoît Mandelbrot

## (1924–2010)

Mathematician, and father of fractal geometry, who described the roughness of nature.

“The financiers and investors of the world are, at the moment, like mariners who heed no weather warnings.”<sup>1</sup> Those words were written by Benoît Mandelbrot four years before the recent financial crisis. Mandelbrot, a mathematician world-famous for his work on fractal geometry, died on 14 October at the age of 85. His financial prescience was a natural outgrowth of his original and penetrating view of the world.

At a time when mathematics focused on lines, planes and spheres, Mandelbrot wrote: “Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.”<sup>2</sup> His life’s work was the creation of ways to describe these objects more accurately. He was able to see and describe the true roughness of the world.

Mandelbrot was born into an educated Jewish family in Poland. As he put it: “I was expected without saying to become a scholar of some sort. Any other activity would have required a specific reason.”<sup>3</sup> In 1936, his family, seeing the rise of Nazi Germany, moved first to Paris and then to a small town in central France. After the fall of France, the threat to Jews from the German occupation was ever present. To survive, Mandelbrot moved often, making his attendance at formal schools irregular. He was briefly a groom and an apprentice toolmaker. At one point, he narrowly escaped deportation and probable death.

During his brief attendance at an advanced school in Lyons, he discovered that he had a remarkable gift for visualizing geometric objects. This gift enabled him to quickly solve difficult algebraic problems in different ways from other students. As the war ended, he returned to Paris and prepared intensively for entry to the grandes écoles — the elite French universities. Despite his uneven schooling, he placed almost at the very top in the examinations and entered the École Polytechnique, then in Paris. He once described to me the confusing emotions he experienced during this sudden transformation from near fugitive to a member of the upcoming technocratic elite, and his belief that his irregular life with its limited schooling had given him the time and the freedom to develop intellectually in his own way.

When Mandelbrot graduated from the Polytechnique, the dominant mathematics in France was pure and abstract. Mandelbrot’s goal, like his background, was different. He



wanted to find order where everyone else saw a lawless mess. He wanted to learn about real, concrete complex problems. He was able to do this with a scholarship at the California Institute of Technology in Pasadena. There he learned about turbulence and was exposed to the molecular biology being developed by Max Delbrück’s group. Returning to Paris, in 1952 he wrote an unorthodox doctoral thesis about the law that governs the frequency with which individual words occur in ordinary language.

In 1958, Mandelbrot returned to the United States with his wife Aliette Kagan, who was to be his devoted companion throughout his life. There he joined IBM’s newly formed Research Division in Yorktown Heights, New York, where his abilities were quickly recognized and where he had almost complete intellectual freedom for more than three decades.

### VARIANCE AND ROUGHNESS

When I first knew Benoît at IBM, he was already modelling the variations (roughness) of stock prices. I remember him telling me that price changes, even over a long period, were concentrated in only a few hectic days of large price swings. He went on to find similar data for the floods of the Nile, cotton prices, wheat prices and interest rates. Real price variations were far rougher and more extreme than those that could emerge from the models then being used. He realized that to obtain realistic results, a model of day-to-day fluctuations having an infinite variance was needed.

Mandelbrot’s thinking about the roughness of natural objects surfaced in a now-famous paper, ‘How Long is the Coast of Britain?

Statistical Self-Similarity and Fractional Dimension’, published in *Science*<sup>4</sup> in 1967. There he first used statistical self-similarity and Hausdorff fractional dimension to describe coastlines in a more accurate way. In subsequent work he extended this approach to describe the shapes of mountains, the branching of rivers and the insides of lungs. In 1975 he coined the term fractal to describe the rough but structured forms he saw all around him. This ever-expanding work appeared in various forms, culminating in his book *The Fractal Geometry of Nature* (1982).

Mandelbrot’s remarkable conclusions often directly contradicted the accepted view. Inevitably, this slowed their acceptance, but he always persisted with an intellectual courage that I greatly admired. In 1974 he became an IBM Fellow, IBM’s highest technical distinction, but outside recognition came more slowly.

Eventually his work took hold, helped both by its intrinsic importance and by the sheer beauty of the pictures that his and others’ work on fractals generated. One fractal, suitably named the Mandelbrot set, became globally recognized, and questions about its properties sparked the interest of many mathematicians. Finally in 1985 he received the Barnard Medal, awarded by the US National Academy of Sciences, and after that came a flood of recognition, honorary degrees, elections to prestigious academies, prizes and the Legion of Honour.

In 1987 he moved to Yale University in New Haven, Connecticut, becoming the Sterling Professor of Mathematics in 1999, and transitioning to emeritus status in 2004. At Yale, he steadily expanded his work and its area of application, surrounded by the fame and recognition his achievements had earned him.

The Wolf Prize citation summarized those achievements well when it said of Mandelbrot “He has changed our view of nature”. ■

**Ralph Gomory** was for many years IBM’s director of research. He is now a research professor at New York University, New York, New York 10012, USA.  
e-mail: gomory@sloan.org

1. Mandelbrot, B. & Hudson, R. L. *The (Mis)behavior of Markets* (Basic Books, 2004).
2. Mandelbrot, B. B. *The Fractal Geometry of Nature* (W. H. Freeman, 1982).
3. Mandelbrot, B. in *One Hundred Reasons to be a Scientist* (ed. Sreenivasan, K. R.) 157 (Abdus Salam International Centre for Theoretical Physics, 2004).
4. Mandelbrot, B. *Science* **156**, 636–638 (1967).

R. RESSMEYER/CORBIS

## Synthetic metabolism goes green

An extension of synthetic biology to a medicinal plant involves the transfer of chlorination equipment from bacteria. This exercise adds implements to the enzymatic toolbox for generating natural products. [SEE LETTER P.461](#)

JOSEPH P. NOEL

Plants offer a wondrous diversity of natural products for the chemist to explore and manipulate<sup>1</sup>. Their genetic, developmental and ecological complexity makes them tough targets, but O'Connor and colleagues (page 461 of this issue<sup>2</sup>) now provide an impressive example of how a plant's biosynthetic pathways can be tuned to fruitful ends.

The organismal complexities posed by plants mean that they have largely been supplanted by microorganisms as a source of natural products; microbial genetics, including the coordinately regulated and sequential arrangement of genes encoding biosynthetic pathways, is much more tractable. By contrast, plant natural products are often built by unknown numbers of enzymes, are encoded by genes lacking the orderly arrangement in microorganisms, and their fate is in part determined by task-oriented cells working in concert for biosynthesis, transport and storage<sup>3</sup>. Regardless of the source, if a natural product is to see the light of day, intervention by chemists is often necessary to create or modify what are arguably some of the most chemically impenetrable scaffolds known<sup>4</sup>.

In their paper, O'Connor and colleagues<sup>2</sup> show that cells of a medicinal plant, *Catharanthus roseus* (Madagascar periwinkle; Fig. 1), can be coaxed into serving as chemical factories by using a combination of genes from microorganisms, enzyme engineering and plant-cell transformation. In this way, they have produced chlorine-containing analogues of natural products called monoterpene indole alkaloids. This alkaloid family includes pharmacologically important compounds that form the backbone of treatments for Hodgkin's disease and acute lymphocytic leukaemia. The biosynthetic installation of a non-natural chlorine atom on particular alkaloid atoms opens up previously inaccessible routes to selective chemical changes<sup>4</sup>. One day, these approaches might be used to alter drug potency and specificity while reducing side effects.

Currently, most metabolic-engineering efforts produce plant compounds by reconstituting, in microbial hosts such as the bacterium *Escherichia coli* or baker's yeast, one or two enzymes known to produce plant secondary



**Figure 1** | *Catharanthus roseus*, more commonly known as the Madagascar periwinkle. This plant is well known to gardeners. But, as O'Connor and colleagues<sup>2</sup> demonstrate, it is also a rewarding subject for the synthetic chemist.

metabolites. Indeed, spectacular successes for high-level production of the antimalarial agent artemisinin, a plant secondary metabolite, have been achieved when two key enzymes of plant metabolic pathways are transformed and optimized in a microbial host<sup>5</sup>. These secondary — or more appropriately 'specialized' — metabolites ensure ecological survival of a plant, and diverge from the ubiquitous primary metabolites required for basal plant physiology. However, little attention has been focused on secondary metabolic engineering in the native plant — particularly in cases where knowledge of the genes encoding the enzymatic toolbox is still incomplete, as is the case for the clinically valuable monoterpene indole alkaloids.

These alkaloids are structurally complex, and are constructed from the common building blocks of the amino acid tryptophan and the ten-carbon terpene geraniol. Although these building blocks are in themselves chemically unassuming, they are transformed by a minimum of 14 enzyme-catalysed steps to form hundreds of monoterpene indole alkaloids<sup>6</sup>. A key step in this metabolic pathway is the removal of the carboxyl group of tryptophan by the *C. roseus* enzyme tryptophan

decarboxylase, producing carbon dioxide and the intermediate tryptamine. Tryptamine then combines with the geraniol-derived terpene product secologanin. Strictosidine, the outcome of this coupling catalysed by the enzyme strictosidine synthase, is converted by a multitude of enzymatic transformations to form the diverse alkaloids of *C. roseus*.

Previously, O'Connor's group<sup>7</sup> showed that synthetic tryptamine analogues containing chlorine atoms and fed to cell cultures of *C. roseus* were easily taken up by the plant cells, and that these chlorinated tryptamines were then incorporated into alkaloid products. The lack of substrate specificity displayed by strictosidine synthase, and by a series of yet uncharacterized downstream enzymes, may seem surprising. But it is increasingly clear that enzymes in general have varying levels of substrate permissiveness and mechanistic flexibility, more commonly referred to as catalytic promiscuity<sup>8</sup>. Indeed, the substrate permissiveness and mechanistic flexibility observed by O'Connor's group in their earlier study<sup>7</sup> is the rule rather than the exception in plant secondary metabolism<sup>1</sup>. Because this earlier research bypassed the need for tryptophan

PHOTOS HORTICULTURAL/PHOTOSHOT



decarboxylase, in the new work<sup>2</sup> O'Connor and colleagues had first to establish that, at least in a test tube, *C. roseus* tryptophan decarboxylase also displays tolerance to chlorine-containing tryptophan substrates.

Having established this substrate permissiveness, the investigators turned to a class of enzyme known as halogenases to add the chlorine atom to one of two carbon atoms of the tryptophan ring. In an ironic twist on conventional metabolic engineering, they employed two genes obtained from soil bacteria that each encode site-specific halide-transfer activity to distinct carbon atoms of the tryptophan ring<sup>9</sup>. One question remained — would the wild-type strictosidine synthase from *C. roseus* accept these two chemically distinct chlorine-bearing tryptamine molecules to afford a larger combinatorial collection of downstream alkaloid products? Although one chlorine-bearing tryptamine was accepted, the other was not. O'Connor and co-workers turned again to earlier results<sup>10</sup> involving structure-based engineering of strictosidine synthase to broaden its substrate selectivity, thus coercing the synthase to accept either of the non-natural tryptamine analogues.

With the necessary biosynthetic toolbox in hand, the authors<sup>2</sup> employed a commonly used soil bacterium, *Agrobacterium rhizogenes*, that can insert foreign genes into plants and plant cell cultures. They generated a type of plant cell culture — known as the 'hairy-root culture' — of *C. roseus* that contained the microbial halogenase genes as well as the mutant form of the *C. roseus* strictosidine synthase. The resulting cultures produced not only the expected chlorinated tryptophan, but also a variety of downstream halogen-containing alkaloids, thereby demonstrating a somewhat surprising level of metabolic promiscuity.

This proof-of-principle study<sup>2</sup> uses a metabolic engineering route to produce collections of modified natural products in cells from overlooked plant hosts that possess the complex enzymatic machinery necessary for these specialized biosyntheses. Many of the crucial enzymes remain unknown and are therefore not genetically accessible for expression in commonly employed microbial hosts. The product yields<sup>2</sup> are modest, but they compare favourably with yields from other test-tube-based reconstitutions of metabolic pathways and from the rudimentary efforts to move plant alkaloid biosynthesis into microbial systems. In addition, the cogent application of catalytic tools amenable to structure-based engineering, and capable of installing a variety of chemical handles on otherwise uncooperative natural products, should expand plant natural-product discovery into the domain of the medicinal chemist.

O'Connor and colleagues' paper<sup>2</sup> provides clear directions for engineering greater catalytic promiscuity into the *C. roseus* tryptophan decarboxylase, thereby alleviating a metabolic

bottleneck in the hairy-root culture system. As such, the work stands as an elegant example of how choosing what seems to be a circuitous experimental route may actually provide a more direct path to success for the synthetic biologist. ■

**Joseph P. Noel** is a Howard Hughes Medical Institute investigator at the Jack H. Skirball Center for Chemical Biology and Proteomics, Salk Institute for Biological Studies, La Jolla, California 92037, USA.  
e-mail: noel@salk.edu

## SPECTROSCOPY

## Clear signals from surfaces

**Nuclear magnetic resonance is a versatile analytical technique, but acquiring well-resolved NMR spectra of chemical surfaces has been hard. The coming of age of a spectral enhancement method should change all that.**

ROBERT G. GRIFFIN

For many decades, nuclear magnetic resonance (NMR) studies of surfaces have promised to provide detailed information about reaction mechanisms involving solid catalysts, but many of the results have been limited by the low signal-to-noise ratio of the experiments. With the recent development of advanced dynamic nuclear polarization (DNP) techniques to enhance NMR sensitivity, however, this situation has changed dramatically. Reporting in the *Journal of the American Chemical Society*, Lesage *et al.*<sup>1</sup> describe an outstanding example of DNP in which the NMR signals of molecules attached to silica surfaces were enhanced approximately 50-fold. Such an increase in sensitivity could revolutionize NMR studies of surfaces.

The physicist Albert Overhauser first proposed<sup>2</sup> the idea of DNP in NMR experiments in 1953, and the concept was demonstrated experimentally by Charles Slichter and co-workers<sup>3</sup> shortly thereafter. NMR spectroscopy involves the use of radio-frequency electromagnetic radiation to excite polarized (aligned) nuclear spins in a magnetic field. But the spin polarizations achieved — and therefore the signal-to-noise ratio of the resulting NMR spectra — are low. However, the spin polarization of electrons in paramagnetic compounds (such as stable free radicals) is hundreds or thousands of times larger than that of nuclei. The DNP technique therefore involves transferring spin polarization from electrons in a paramagnetic compound to the nuclei in a surface sample. This is accomplished

1. Austin, M. B., O'Maille, P. E. & Noel, J. P. *Nature Chem. Biol.* **4**, 217–222 (2008).
2. Rungtaphan, W., Qu, X. & O'Connor, S. E. *Nature* **468**, 461–464 (2010).
3. Ziegler, J. & Facchini, P. J. *Annu. Rev. Plant Biol.* **59**, 735–769 (2008).
4. Deb Roy, A., Gröschow, S., Cairns, N. & Goss, R. J. M. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010).
5. Ro, D.-K. *et al.* *Nature* **440**, 940–943 (2006).
6. O'Connor, S. E. & Maresh, J. J. *Nat. Prod. Rep.* **23**, 532–547 (2006).
7. Bernhardt, P., McCoy, E. & O'Connor, S. E. *Chem. Biol.* **14**, 888–897 (2007).
8. Tawfik, D. S. *Nature Chem. Biol.* **6**, 692–696 (2010).
9. Ryan, K. S. & Drennan, C. L. *Chem. Biol.* **16**, 351–364 (2009).
10. Loris, E. A. *et al.* *Chem. Biol.* **14**, 979–985 (2007).

by irradiating the electron paramagnetic resonance (EPR) spectrum, the electronic analogue of the NMR spectrum, with microwaves, thereby exciting electron–nucleus transitions and transferring polarization.

In the 1980s, DNP was combined with magic-angle spinning (MAS), an NMR technique used to obtain high-resolution spectra of solids, to enhance the sensitivity of NMR for studying polymers and other materials<sup>4–6</sup>. These were 'low-field' experiments — they used a relatively low magnetic field (1.5 tesla), low radio frequencies (60 megahertz for NMR of <sup>1</sup>H nuclei) and low microwave frequencies (40 GHz for EPR). But at that time, MAS was rapidly moving towards using higher magnetic fields (5–20 T) and radio frequencies (200–850 MHz for <sup>1</sup>H), which offer greater resolution and sensitivity. To obtain large signal enhancements from DNP in such high-field experiments requires microwave sources operating at 130–600 GHz. Such sources were not readily available at the time, and so DNP–MAS failed to take off as a solid-state analytical technique. DNP therefore resumed its former position as an interesting intellectual curiosity.

The 1990s witnessed the development of several pieces of instrumentation that altered this landscape considerably. For instance, 1993 saw the introduction of a class of microwave oscillators known as gyrotrons<sup>7</sup> for DNP. The first gyrotrons<sup>8–10</sup> provided continuous microwave power at 140 and 250 GHz, but more recent devices<sup>11,12</sup> do so at up to 460 GHz (which could be used for <sup>1</sup>H DNP–NMR at 700 MHz). Microwave sources corresponding to the highest available NMR frequencies



## 50 Years Ago

A recent broadsheet issued by Political and Economic Planning, entitled "The Growing Economy—Britain, Western Germany and France", discusses the main factors influencing the development of industrial production ... These findings are to be published in a forthcoming report which, assuming that economic growth is a proper object of policy, considers the conditions for achieving a rate of growth comparable with that of other Western industrial nations ... In contrast to British policy, French and German economic policies have had marked success in stimulating economic growth, and the main conclusion of the broadsheet is that, in view of Britain's record since the War, priority must be given to the task of increasing the rate of growth.

From *Nature* 19 November 1960

## 100 Years Ago

*The Roosevelts in Africa* — The book under review is not without its defects and incongruities, and the expedition of which it is the record has received heavy censure from a good many people interested in the preservation of the world's fauna. Theodore Roosevelt, its author, has the defects of his qualities ... In the first place, Mr. Roosevelt has not had sufficient leisure in which to do himself justice as the writer of a book on real natural history. Being a poor man when he left the Presidency, he was obliged, to a great extent, to pay the expenses of his very costly expedition by writing an account of it to be published week by week by the newspapers, a full diary, so to speak, of the day's events. Then, taking advantage of a brief rest at Khartum, he puts this diary together in book form, and has barely time to glance at the proofs before leaving England for the States in June.

From *Nature* 17 November 1910

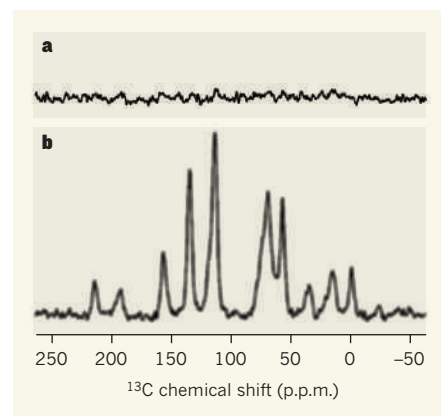
(1,000 MHz) are now on the horizon. The fact that DNP functions optimally at low temperatures has also necessitated the development of a new generation of cryogenic MAS probes that operate at temperatures of 90 kelvin and below<sup>13,14</sup>.

Another major development was the discovery of innovative polarizing agents. In the first 50 years of DNP experiments, researchers used single electrons (mainly from organic free radicals) as polarizing agents. These act through a mechanism known as the solid effect, which involves two spins — the electron's spin and the spin of the nucleus to be polarized. But a more efficient process involving three spins was demonstrated in 2004, with the development of biradical polarizing agents<sup>15</sup>. Compared with experiments in the absence of DNP, 250-fold signal enhancements have been observed<sup>16</sup> using such biradicals. Currently, the favourite polarizing agent is a water-soluble biradical known as TOTAPOL, which typically yields 170-fold enhancements in model systems and in proteins.

Lesage and colleagues' exciting NMR experiments<sup>1</sup> bring together all of the state-of-the-art developments in DNP-enhanced NMR. They used an NMR spectrometer equipped with a 263-GHz gyrotron, operating at 90 K with TOTAPOL as the polarizing agent, to enhance the spectra of organic groups covalently attached to the surface of porous silica. Surface-modified silica is commonly used in many applications, and is a good case study for various other systems whose surface chemistry is ripe for DNP–NMR studies. The authors observed a 50-fold enhancement of the carbon-13 NMR signals for the silica-bound groups (Fig. 1), which allowed the acquisition of <sup>13</sup>C spectra in approximately 30 minutes. In the absence of DNP, these experiments would have taken about 70 days. Refinements to the technique could eventually yield approximately 500-fold signal enhancements.

The mechanism by which the enhanced NMR signals<sup>1</sup> are most frequently generated involves polarizing the solvent in which the silica material is suspended. This polarization is transferred to the surface of the material (and probably deeper than that) through spin diffusion processes. Other experiments<sup>18,19</sup> on nanocrystals and membranes have shown that polarization can be transferred over distances of around 1,000 ångströms, so it may be possible to examine the structure of a material not only at the surface, but also in the layers immediately below. This is one of the many possibilities envisaged by Lesage and colleagues<sup>1</sup>.

The authors' experiments represent a huge step forward in the study of reactions that occur at the surfaces of bulk solids, including many scientifically and industrially important reactions on solid catalysts. The application of DNP–NMR to such systems will undoubtedly stimulate many new avenues of research. More generally, the authors' work<sup>1</sup> represents another



**Figure 1 | Enhanced NMR signals from surfaces.** Lesage *et al.*<sup>1</sup> report an impressive signal enhancement for NMR spectra of organic groups attached to a silica surface using a technique known as dynamic nuclear polarization (DNP). **a**, Without DNP, the authors observed no sharp peaks in the carbon-13 spectrum of the surface groups. **b**, With DNP, they obtained a 50-fold improvement in the NMR signal. The spectra in **a** and **b** are shown at the same scale. Peaks at different chemical shifts correspond to different carbon atoms in the surface groups. Spectral data are from ref. 1; p.p.m. denotes parts per million.

example of the application of DNP to heterogeneous systems (those that involve more than one phase of matter), whose structures are often difficult to determine. ■

**Robert G. Griffin** is in the Department of Chemistry and the Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

e-mail: rgg@mit.edu

1. Lesage, A. *et al.* *J. Am. Chem. Soc.* **132**, 15459–15461 (2010).
2. Overhauser, A. W. *Phys. Rev.* **92**, 411–415 (1953).
3. Carver, T. R. & Slichter, C. P. *Phys. Rev.* **92**, 212–213 (1953).
4. Wind, R. A., Duijvestijn, M. J., van der Lugt, C., Manenschijn, A. & Vriend, J. *Prog. Nucl. Magn. Reson. Spectrosc.* **17**, 33–67 (1985).
5. Afeworki, M. & Schaefer, J. *Macromolecules* **25**, 4092–4096 (1992).
6. Singel, D. J., Seidel, H., Kendrick, R. D. & Yannoni, C. S. *J. Magn. Reson.* **81**, 145–161 (1989).
7. Felch, K. L. *et al.* *Proc. IEEE* **87**, 752–781 (1999).
8. Becerra, L. R., Gerfen, G. J., Temkin, R. J., Singel, D. J. & Griffin, R. G. *Phys. Rev. Lett.* **71**, 3561–3564 (1993).
9. Gerfen, G. J. *et al.* *J. Chem. Phys.* **102**, 9494–9497 (1995).
10. Bajaj, V. S. *et al.* *J. Magn. Reson.* **189**, 251–279 (2007).
11. Matsuki, Y. *et al.* *Phys. Chem. Chem. Phys.* **12**, 5799–5803 (2010).
12. Torrezan, A. C. *et al.* *IEEE Trans. Plasma Sci.* **38**, 1150–1159 (2010).
13. Barnes, A. B. *et al.* *J. Magn. Reson.* **198**, 261–270 (2009).
14. Thurber, K. R. & Tycko, R. *J. Magn. Reson.* **195**, 179–186 (2008).
15. Hu, K.-N., Yu, H., Swager, T. M. & Griffin, R. G. *J. Am. Chem. Soc.* **126**, 10844–10845 (2004).
16. Matsuki, Y. *et al.* *Angew. Chem. Int. Edn* **48**, 4996–5000 (2009).
17. Song, C., Hu, K.-N., Joo, C.-G., Swager, T. M. & Griffin, R. G. *J. Am. Chem. Soc.* **128**, 11385–11390 (2006).
18. van der Wel, P. C. A., Hu, K.-N., Lewandowski, J. & Griffin, R. G. *J. Am. Chem. Soc.* **128**, 10840–10846 (2006).
19. Barnes, A. B. *et al.* *Phys. Chem. Chem. Phys.* **12**, 5861–5867 (2010).



## NEUROSCIENCE

# Excessive mobility interrupted

**Mobile DNA sequences called L1 contribute to the brain's genetic heterogeneity and may affect neuron function. The protein MeCP2, which is mutated in Rett syndrome, seems to regulate the activity of these genomic elements. [SEE LETTER P.443](#)**

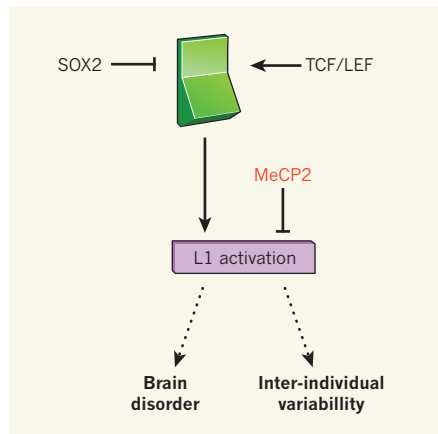
LORENZ STUDER

In his struggle to rationalize the laws of quantum mechanics, Albert Einstein once wrote that God “does not play dice”. A similar struggle comes to mind when trying to understand the biological consequences of several reports from the labs of Gage and Muotri, including one appearing on page 443 of this issue (Muotri *et al.*<sup>1</sup>). These studies offer evidence for the occurrence of quasi-random genetic changes in neurons during development. The work raises provocative questions about whether ‘playing dice’ during brain development contributes to the differences that make each of us unique, and how such changes may be related to brain disorders.

The immune system is well known for using genetic recombination and random mutations<sup>2</sup> to adapt antibody defences against invaders. The question is, does the nervous system use a similar approach to create neuronal diversity? For example, each olfactory neuron expresses only one of more than 1,000 possible olfactory-receptor genes. Permanent genetic changes are unlikely to underlie selection of olfactory-receptor genes for expression: mice generated from the genetic material of a neuron expressing a single olfactory receptor — by the technique of somatic-cell nuclear transfer — re-establish the full complement of olfactory receptors<sup>3,4</sup>. Such evidence, however, does not rule out the possibility of other permanent genetic changes occurring within the neuronal lineage.

Transposons, or ‘jumping genes’<sup>5</sup>, can mediate one such mechanism, inducing changes in the DNA. In the human genome, the most common class of transposons is retrotransposons, which have the ability to amplify themselves. For instance, one type of retrotransposon, known as LINE-1 or L1, constitutes a staggering 17% of the human genome. However, only a small fraction of L1 is functionally intact and active, and the role of these retrotransposons in the human genome remains mysterious.

It has been proposed<sup>6,7</sup> that L1 is involved in genome evolution, and there is evidence<sup>8</sup> that retrotransposition can induce genetic changes responsible for human disease. Until recently, L1 activity was thought to be confined to the earliest stages of embryonic development and the germ-cell lineage. But this picture started



**Figure 1 | Regulation of L1 expression.** During neuronal development, the expression of L1 retrotransposons is transiently activated when neuronal precursor cells undergo a transcriptional switch from expressing SOX2, which represses L1 expression, to expressing TCF/LEF, which supports it. Muotri *et al.*<sup>1</sup> find that another protein, MeCP2, also regulates L1 expression, repressing its promoter activity and so its retrotransposition rates. Such genetic changes could contribute to brain disorders and variability among individuals.

to crumble a few years ago, when the Gage lab demonstrated<sup>9</sup> active L1 retrotransposition in neuronal precursor cells. Subsequent mechanistic studies identified<sup>10</sup> dual binding sites in the promoter region of L1 for the transcriptional regulators SOX2 and TCF/LEF, which are responsible for changes in the associated chromatin (DNA–protein complexes), and so for the switch between repression and activation of L1 expression (Fig. 1).

The intricate regulation of L1 activity was the starting point for the next chapter of this remarkable story, as told by Muotri and colleagues<sup>1</sup>. Previous work in cell lines<sup>11</sup> had suggested that the protein MeCP2 can recruit the enzyme HDAC1 and that the two contribute to L1 repression. A role for MeCP2 in regulating L1 activity was particularly intriguing, because the gene encoding this protein is mutated in Rett syndrome (RTT) — an X-chromosome-linked disorder specific to the nervous system that is a leading genetic cause of mental retardation in girls.

Muotri *et al.* now use a broad armamentarium of approaches to investigate the link

between MeCP2 and L1 retrotransposition. They present compelling *in vitro* and *in vivo* evidence that both promoter activity and L1 retrotransposition rates are significantly increased in neuronal precursors of mice that lack the *Mecp2* gene. Indeed, the authors' detailed imaging studies demonstrate clear differences in L1 retrotransposition rates between normal mice and those lacking *Mecp2* across several regions of the adult brain.

Muotri and co-workers also investigate the involvement of MeCP2 in regulating L1 retrotransposition in the cells of patients with RTT. For this, they reprogram the patients' skin fibroblast cells into patient-specific induced pluripotent stem cells (iPSCs), and detect increased L1 retrotransposition rates in neurons derived from the iPSCs of patients with RTT compared with neurons from control iPSCs. The authors further corroborate these data with post-mortem work, finding increased levels of L1 DNA content in the brain, as opposed to matched heart-tissue samples, of patients with RTT, and in brain samples from patients compared with those from age-matched controls.

These findings<sup>1</sup> provide intriguing evidence for the modulation of L1 activity by MeCP2 during development of the nervous system. Is such modulation causally involved in RTT pathology? On the basis of the current knowledge of the disease, probably not.

There is no obvious correlation between the timing of L1 activity during embryonic development and the delayed disease onset in patients with RTT, which typically occurs 1–2 years after birth. It is also difficult to imagine how an increased rate of quasi-random genetic changes, even if biased towards genes expressed in neurons<sup>9</sup>, can lead to the highly reproducible disease symptoms observed in RTT. What's more, in *Mecp2*-mutant mice, a striking recovery from RTT symptoms occurs on re-expression of *Mecp2*, even in mature animals — a strategy that clearly does not affect early retrotransposition events. Therefore, changes in L1 activity might not cause RTT, but may contribute to variability among patients — beyond the well-known differences in mutation type and X-chromosome inactivation status.

Another intriguing issue relates to the use of RTT iPSCs in modelling human disease. After reprogramming to iPSCs, the inactive X chromosome of differentiated cells does not become active<sup>12</sup>. Therefore, all differentiated cells generated from a given RTT iPSC line are expected to show identical inactivation patterns of either their normal X chromosome or their *MECP2*-mutant X chromosome, rather than the random inactivation pattern observed in patients. The increased L1 activity in neuronal precursors derived from RTT iPSCs indicates that these cell lines showed inactivation of the normal X chromosome and may, therefore, model a more severe form of RTT.

The holy grail for defining the functional impact of L1 activation in neuronal development would be a method that can selectively switch on and off retrotransposition events using genetic or pharmacological tools. Given the sheer number and genetic complexity of transposable elements, this remains a daunting task, although targeting the reverse transcriptase enzyme or other mission-critical determinants of L1 activity may represent a potentially tractable approach. Until then, we are left to wonder whether playing dice during

development of the central nervous system is indeed part of what makes each of us unique in both health and disease. ■

**Lorenz Studer** is in the Center for Stem Cell Biology, Developmental Biology Program, Sloan-Kettering Institute for Cancer Research, New York, New York 10065, USA.  
e-mail: studerl@mskcc.org

1. Muotri, A. R. *et al.* *Nature* **468**, 443–446 (2010).
2. Tōnegawa, S. *Nature* **302**, 575–581 (1983).
3. Eggen, K. *et al.* *Nature* **428**, 44–49 (2004).

4. Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. *Nature* **428**, 393–399 (2004).
5. McClintock, B. *Proc. Natl Acad. Sci. USA* **36**, 344–355 (1950).
6. Xing, J. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 17608–17613 (2006).
7. Cordaux, R. & Batzer, M. A. *Nature Rev. Genet.* **10**, 691–703 (2009).
8. Kazazian, H. H. Jr *et al.* *Nature* **332**, 164–166 (1988).
9. Muotri, A. R. *et al.* *Nature* **435**, 903–910 (2005).
10. Kuwabara, T. *et al.* *Nature Neurosci.* **12**, 1097–1105 (2009).
11. Yu, F., Zingler, N., Schumann, G. & Strätling, W. H. *Nucleic Acids Res.* **29**, 4493–4501 (2001).
12. Tchieu, J. *et al.* *Cell Stem Cell* **7**, 329–342 (2010).

## QUANTUM PHYSICS

# Entangled quartet

Quantum physics is known for its counter-intuitive principles. One such principle — that a single photon can be in as many as four places at the same time — has now been demonstrated. [SEE LETTER P.412](#)

VLADAN VULETIC

When light is shone through two closely separated slits and onto a distant screen, a periodic light pattern emerges as a result of interference between the light waves emanating from the two slits. Where quantum physics is concerned, some of the deepest mysteries — or, in the opinion of the iconic Richard Feynman, the only mystery — arise when that experiment is performed not with strong classical light waves but with a single particle. Although indivisible, a single particle also produces an interference pattern, so it must have passed simultaneously through both slits.

Building on recent advances<sup>1</sup> enabling the storage of single photons in atomic gases, Choi *et al.*<sup>2</sup> (page 412 of this issue) investigate what happens to interference when light is stored

simultaneously in as many as four spatially distinct atomic clouds. The authors demonstrate quantum correlations (entanglement) in this composite matter–light system, and study how entanglement ultimately fades away to leave only classical correlations.

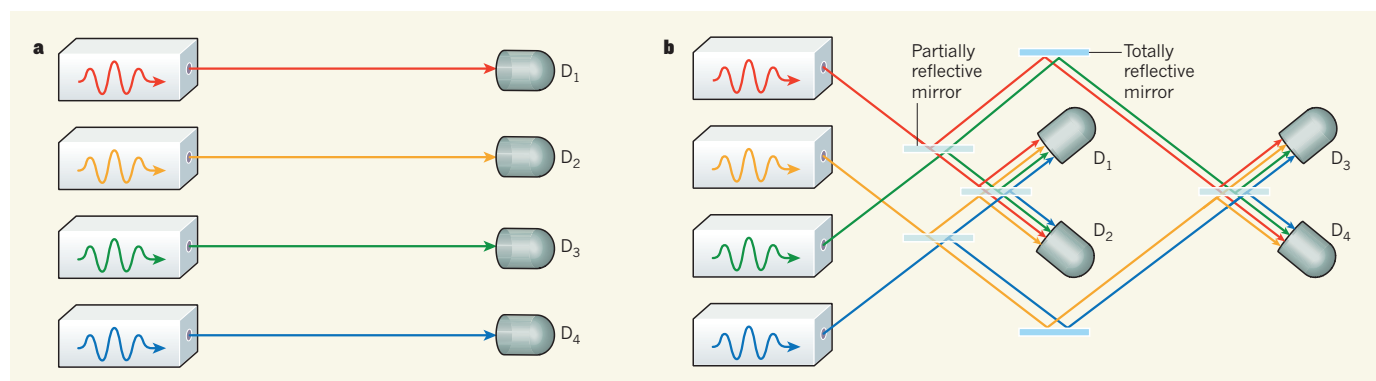
Classical correlations can arise in situations in which there is limited knowledge of a system. For instance, if we know only that one coin (or photon) has been hidden in one of four boxes, then detecting the coin in one box would instantaneously tell us that the other three boxes are empty — even if they were separated from each other by light years. It is hardly surprising that such ‘particle-type’ detection (Fig. 1a) can reveal classical correlations between the numbers of coins found in the different boxes if the total number of coins is known a priori.

Classical correlations can also arise between

multiple light waves that are combined on partially reflecting mirrors before detection, such that the origin of the detected light is unknown (wave-type detection). For example, if identical light waves have been stored simultaneously in all four boxes — or, for that matter, coins sufficiently small to display quantum, wave-like character — and the light emerging from the boxes is combined through a series of partially reflecting and totally reflective mirrors before reaching four detectors (Fig. 1b), then the outputs of the detectors would vary as the path length between each box and the corresponding first mirror is varied.

In a classical world, something is either a particle or a wave, so a physical system will exhibit correlations either in the particle-type or wave-type detection set-up — but not in both. However, in the quantum world that we live in, it is possible to place, for example, a single photon simultaneously in all boxes such that correlations are observed in both detection set-ups. And this is exactly what Choi *et al.*<sup>2</sup> have done in their experiment.

Choi and colleagues used four atomic ensembles as the storage boxes. Such systems not only can hold the photon, but also can act as highly directional light emitters that can be triggered on demand through the application of a laser pulse<sup>1,3</sup>. The authors measured correlations between the different boxes, either in



**Figure 1 | Particle-type versus wave-type measurements.** Choi *et al.*<sup>2</sup> have measured quantum entanglement in a composite matter–light system by combining results from particle-type and wave-type measurements. The matter component of the system consists of four atomic ensembles (illustrated by the boxes) and the light part is a single photon (waveform). **a**, In the particle-type set-up, a photon stored in one box can reach only one detector ( $D_1$ ,  $D_2$ ,  $D_3$

or  $D_4$ ). **b**, In the wave-type measurement, the photon is placed simultaneously in all four boxes and the light emerging from the boxes is combined through an arrangement of partially reflecting and totally reflecting mirrors such that light from any box can reach any detector. The colours and multiple waveforms are for illustration of the photon path only; the light in all four boxes is identical, has the same wavelength, and contains only one photon in total.



the particle-type detection set-up (Fig. 1a) or in the wave-type set-up (Fig. 1b). From the combination of these measurements, they extracted the degree of entanglement of the light shared between the four boxes. Using a method previously developed<sup>4</sup> for a single photon travelling simultaneously along four possible paths, they identified quantitative criteria, involving combinations of particle-type and wave-type detection results, that allowed them to distinguish among entanglement between all four boxes, or three, or just two of them. In the presence of noise and other imperfections, they observed a gradual transition from four-party entanglement to no entanglement.

Although entanglement among more than four parties has been observed (the current record is for a system of 14 ions<sup>5</sup>, and entanglement has been inferred among 100 atoms<sup>6</sup>),

Choi and colleagues' system<sup>2</sup> is special because the entanglement can be efficiently mapped on demand from a material system onto a light field. Atomic ensembles such as those used by the authors have already reached light-storage times of milliseconds at the single-photon level<sup>7,8</sup>. If those storage times can be extended to seconds, and some other technical performance parameters improved, such sources will have a variety of potential applications in secure quantum communication over long distances<sup>1</sup>. The ensembles could then be used to build quantum networks over which quantum information can be distributed.

The astute reader may wonder how it is that quantum correlations can be observed with a single photon given that any correlation requires more than one system. The controversy about this issue can be resolved<sup>9</sup>

by viewing the four boxes as the systems that exhibit correlations (in photon number), rather than considering a single photon with qualms about its parent box. ■

**Vladan Vuletic** is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.  
e-mail: vuletic@mit.edu

1. Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. *Nature* **414**, 413–418 (2001).
2. Choi, K. S., Goban, A., Papp, S. B., van Enk, S. J. & Kimble, H. J. *Nature* **468**, 412–416 (2010).
3. Vuletic, V. *Nature Phys.* **2**, 801–802 (2006).
4. Papp, S. B. *et al. Science* **324**, 764–768 (2009).
5. Monz, T. *et al. Preprint at* <http://arxiv.org/abs/1009.6126> (2010).
6. Gross, C. *et al. Nature* **464**, 1165–1169 (2010).
7. Zhao, R. *et al. Nature Phys.* **5**, 100–104 (2009).
8. Zhao, B. *et al. Nature Phys.* **5**, 95–99 (2009).
9. van Enk, S. J. *Phys. Rev. A* **72**, 064306 (2005).

## FISHERIES

# Measuring biodiversity in marine ecosystems

**The use of catch data to determine indicators of biodiversity such as 'mean trophic level' does not adequately measure ecosystem changes induced by fishing. Improved ways to assess those changes are required. SEE LETTER P.431**

JOSEPH E. POWERS

**A**ccurate indicators of biodiversity are essential for managing exploited marine ecosystems. The currently most widely adopted indicator is the 'mean trophic level' of catches, the position of a specific species in the food chain (trophic level) averaged over all the species in the catch. Declines in catch mean trophic levels have been interpreted as showing shifts in ecosystem diversity from high-trophic-level predators to lower-trophic-level species. But are indicators based on catch data accurately depicting what is happening to an ecosystem? This question has now been addressed by Branch and co-workers on page 431 of this issue<sup>1</sup>.

Catch databases from marine fisheries are a reflection of economic, biological, ecological and technological factors. As a result, some species are unduly emphasized in the catches, distorting their true occurrence in the ecosystem. Additionally, catch databases, or more correctly 'reported catches', might not reflect the full extent of exploitation. Discarded bycatch, recreational fisheries and rare species are difficult to monitor and are therefore often not fully represented in the data. Finally, the databases themselves are often organized around political jurisdictions and do not necessarily encompass the entire ecosystem. Nevertheless,

catch databases are easily accessible and have relatively comprehensive species composition. So, despite the drawbacks, they remain attractive for formulating diversity indicators such as indices of mean trophic level.

Branch *et al.*<sup>1</sup> examined how useful these databases really are. They did this by comparing the mean trophic level of catches with the mean trophic level of ecosystems (mean trophic level weighted by the estimated true abundance of species in the ecosystem), using two avenues of research.

First, they collated 25 existing and well-documented marine-ecosystem models, representing regions in the Northern and Southern Hemispheres, over a wide range of latitudes. For each model, components encompassing the existing fisheries of the region had already been incorporated. Time series of catch and abundance were projected under four fishing scenarios: 'fishing down', in which higher trophic levels were fished to depletion followed by the advent of fishing on lower trophic levels; 'fishing through', in which there was an expansion of fishing from some higher trophic species to other higher and lower trophic species; fishing 'based on availability', in which those species that were most abundant and accessible were exploited first, followed by expansion to less available and abundant species; and 'increase to overfishing', in which exploitation

rates of all species gradually increased until they were overfished. The simulation projections were used to compute catch- and abundance-weighted trophic indices and compare their time series.

Branch and colleagues' second method was to compare catch- and abundance-weighted mean trophic levels for individual ecosystems. They used relative abundance from trawl surveys from 29 ecosystems, representing regions in the Northern and Southern Hemispheres, five continents and various latitudes, to calculate ecosystem (abundance-weighted) trophic indices. Additionally, estimates of absolute abundance from a database of 242 single-species stock assessments were also used to compute abundance-weighted trophic indices.

The results showed an inconsistent relationship between catch- and abundance-weighted trophic indices. In other words, catch-weighted trophic indices are not generally indicative of the changes in trophic level of the ecosystem. For example, simulated trophic indices from the ecosystem models, as depicted in the top two rows of the authors' Figure 1 (page 431), showed that in some cases the decline in ecosystem mean trophic level (blue lines) was more rapid than that of the catch mean trophic level (red lines), particularly when 'fishing down' occurred. In other cases, the change in mean trophic level of either the catch or the ecosystem was hardly noticeable, yet many species were depleted.

When the individual ecosystems were examined, almost half of the comparisons between catch mean trophic level and ecosystem mean trophic level from trawl or stock-assessment data were found to be negatively correlated. In particular, the relationship between catch and ecosystem trophic level tends to break down when fishing is not distributed across all portions of the ecosystem. On the face of it, then, the way forward is to use abundance-weighted rather than catch-weighted indices.

However, abundance databases have their limitations, too. Although stock-assessment estimates of abundance are considered to provide the best available data<sup>2,3</sup>, the suite of species for which such assessments are done are limited, being driven by economic and management considerations rather than ecological factors. Trawl survey data provide relative abundance estimates that are skewed by differential susceptibilities of the species and sizes to the sampling gear. Additionally, surveys are not normally designed to sample top predators. The results of Branch *et al.* highlight the need to expand research to estimate abundance through stock assessments of a broader range

of species and more extensive trawl surveys.

But is there still some utility in using catch-weighted mean trophic levels? Perhaps so. Branch and colleagues' results<sup>1</sup> suggest conditions in which they might be useful (for example, to indicate major shifts in exploitation patterns). Additionally, catch-weighted trophic level might be used as a 'policy-triggering' tool rather than as a monitoring index — that is, a major change in catch mean trophic level would trigger more detailed research and/or more precautionary management strategies. Indeed, it can be argued that this is exactly how catch-weighted mean trophic levels have been used previously, in that they have provoked

consideration of broad ecosystem policy issues. However, further simulation research is needed to evaluate which management actions are most effective for specific ecosystems. Branch *et al.* have provided the basis for doing that. ■

**Joseph E. Powers** is in the Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA.  
e-mail: jepowers@lsu.edu

1. Branch, T. A. *et al.* *Nature* **468**, 431–435 (2010).
2. Polacheck, T. *Mar. Policy* **30**, 470–482 (2006).
3. Worm, B. *et al.* *Science* **325**, 578–585 (2009).

## REPRODUCTIVE AGEING

# Of worms and women

**In roundworms, age-related decline in egg quality is regulated by specific humoral signalling pathways. If similar mechanisms operate in mammals, these findings may suggest ways to delay reproductive ageing in women.**

KEVIN FLURKEY & DAVID E. HARRISON

Female mammals are not alone in experiencing an age-related increase in birth defects and decline in fertility; the roundworm *Caenorhabditis elegans* faces similar reproductive challenges in mid-adulthood. Writing in *Cell*, Luo *et al.*<sup>1</sup> report that, in *C. elegans*, the age-related decline in oocyte (egg) quality and increase in chromosomal abnormalities are regulated by evolutionarily conserved signal-transduction pathways. If this senescence mechanism is also conserved, age-related decline in the quality of mammalian oocytes may not be, as is commonly thought, simply due to the old age of these cells or the diminishing size of the ovarian follicle pool; it may also be influenced by molecular signalling cascades.

Previous work in *C. elegans* showed<sup>2</sup> that a mutation that reduces the function of *daf-2* — a gene involved in an insulin/IGF-I-like signalling pathway — delays reproductive senescence. Moreover, in an earlier study<sup>3</sup>, Luo and colleagues showed that reproductive lifespan is extended by mutations that decrease the activity of the TGF- $\beta$  Sma/Mab signalling pathway, which regulates cell growth, body size and the development of male traits.

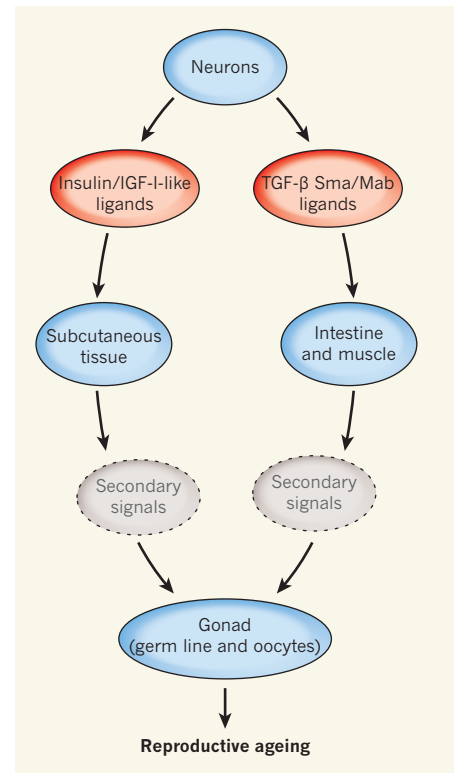
Confirming and extending these findings, Luo *et al.*<sup>1</sup> now show that decreasing activity in both of these pathways increases reproductive lifespan by delaying age-specific reductions in germline cell numbers, oocyte fertilizability and embryo hatching, as well as by diminishing the age-related increase in chromosomal abnormalities. Using *C. elegans* stocks with pathway-specific and tissue-specific mutations

in components of the insulin/IGF-I or TGF- $\beta$  Sma/Mab pathways, the authors show that these signalling cascades act at distal sites to affect germline function. In fact, they propose a model to describe such neuroendocrine regulation of reproductive senescence (Fig. 1).

These ideas could be of clinical relevance owing to similarities in oocyte development between *C. elegans* and humans. In both species, oocyte development is temporarily halted at the prophase I stage of meiotic cell division, when chromosomal abnormalities most frequently occur. What's more, chromosomal abnormalities — including aneuploidies that result from chromosome non-disjunction — are the main defect in human embryos from ageing mothers<sup>4</sup>, and rates of chromosome non-disjunction also increase with age in *C. elegans*.

Luo *et al.*<sup>1</sup> observe other aspects of diminished oocyte quality with reproductive ageing in *C. elegans* that are similar to those previously reported in older women. For instance, the authors' transcriptional analyses of worms with mutations in TGF- $\beta$  signalling indicate that numerous molecular mechanisms that have a bearing on age-related diminished oocyte quality are influenced by this pathway, and that many of these mechanisms are shared between *C. elegans* and humans.

Two main species differences, however, temper the understandable enthusiasm that Luo and colleagues express for the possibility of translational application of their work to humans. First, the reproductive system of female mammals is more complex than that of the roundworm, with ageing involving both neuroendocrine and oocyte defects<sup>5</sup>. Second, the progressive shrinkage that occurs in the



**Figure 1 | Neuroendocrine regulation of reproductive ageing<sup>1</sup>.** Neurons secrete ligands that act on cells of the intestine and muscle tissue (the TGF- $\beta$  Sma/Mab ligands) or the subcutaneous tissue (insulin/IGF-I ligands) to generate as-yet-unidentified secondary signals. The secondary signals affect germline senescence by altering these cells' morphology, diminishing their proliferation, reducing oocyte quality and increasing chromosomal abnormalities. Consequently, embryonic viability declines and infertility increases.

pool of ovarian follicles during mammalian ageing has no parallel in roundworms.

Indeed, mammals stop producing oocytes even before birth, whereas roundworms continue to produce them throughout their reproductive lives. Theories of mammalian reproductive ageing posit that the age-related decline in oocyte number is the primary factor driving decline in fertility, with the associated deterioration in oocyte quality and increased



risk of birth defects being consequences of a suboptimal size of the oocyte pool. These theories are largely supported by studies in rats<sup>6</sup> showing that reproductive ageing is more closely related to the size of the oocyte pool than to oocyte age.

Nevertheless, Luo and colleagues' work<sup>1</sup> clearly illustrates that neuroendocrine cascades mediate oocyte ageing and embryo survival in *C. elegans*. Because such signalling has clear parallels in humans, straightforward clinical interventions that diminish the effects of insulin/IGF-I or TGF- $\beta$  signalling, or both, may be feasible. However, this depends on how closely neuroendocrine-directed reproductive ageing in *C. elegans* models mammalian reproductive ageing.

So far, the results are inconclusive. The diminished fecundity seen in young adult *C. elegans* as a result of a reduction in insulin/IGF-I signalling<sup>2</sup> may be comparable to the modest reproductive impairments of mouse models with diminished circulating levels of IGF-I<sup>7</sup>. Compared with age-matched controls, the pool of primordial oocytes in these mice is larger<sup>7</sup>, but no associated increase in reproductive lifespan has been reported. Moreover, our unpublished data indicate that, in 'little' mice (C57BL/6J-*Ghrhr*<sup>fl</sup>), a 90% reduction in circulating IGF-I levels has no effect on litter size or females' reproductive lifespans. In *C. elegans*, only one of the numerous reduction-of-function mutations in the insulin/IGF-I signalling pathway increases reproductive lifespan<sup>8</sup>.

The contrasting effects of various insulin/

IGF-I signalling mutants show that subtle differences in perturbations of insulin/IGF-I signalling greatly affect reproductive outcomes in both species. The effects of diminished TGF- $\beta$  signalling on reproductive lifespan in mammals has not been evaluated, and should be. In addition, it is essential to identify the secondary signals that act directly on the *C. elegans* germ line and oocytes to diminish their quality (Fig. 1).

Does Luo and co-workers' paper<sup>1</sup> have a take-home message for women concerned about getting pregnant later in life and giving birth to healthy babies? No. However, this study offers specific hypotheses that can be tested in mammalian systems. It thus opens the door to the possibility of improving oocyte quality during the period of reproductive decline by reducing the effects of insulin/IGF-I or TGF- $\beta$  signalling pathways, or both. ■

**Kevin Flurkey and David E. Harrison**  
are in *The Jackson Laboratory, Bar Harbor, Maine 04609, USA.*  
e-mail: david.harrison@jax.org

1. Luo, S., Kleemann, G. A., Ashraf, J. M., Shaw, W. M. & Murphy, C. T. *Cell* **143**, 299–312 (2010).
2. Hughes, S. E., Evason, K., Xiong, C. & Kornfeld, K. *PLoS Genet.* **3**, e25 (2007).
3. Luo, S., Shaw, W. M., Ashraf, J. & Murphy, C. T. *PLoS Genet.* **5**, e1000789 (2009).
4. te Velde, E. R. & Pearson, P. L. *Hum. Reprod. Update* **8**, 141–154 (2002).
5. Downs, J. L. & Wise, P. M. *Mol. Cell. Endocrinol.* **299**, 32–38 (2009).
6. Meredith, S., Dudenhoeffer, G., Butcher, R. L., Lerner, S. P. & Walls, T. *Biol. Reprod.* **47**, 162–168 (1992).
7. Slot, K. A. et al. *Reproduction* **131**, 525–532 (2006).
8. Tatar, M. *Ann. NY Acad. Sci.* **1204**, 149–155 (2010).

## MATERIALS CHEMISTRY

# Thin films with a hidden twist

**Many naturally occurring substances have a 'handedness' that enables them to interact highly specifically with matter or light. The helical features responsible for this can now be replicated in solid, porous films. [SEE LETTER P.422](#)**

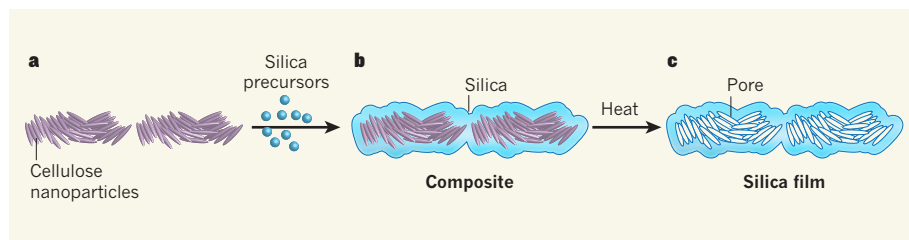
ANDREAS STEIN

**P**orous silica can be made to replicate the helical arrangement of nanocrystalline cellulose derived from wood pulp if, as Shopsowitz et al.<sup>1</sup> demonstrate on page 422 of this issue, the silica is formed around a cellulose scaffold at a specific pH. Using a simple, scalable method, the researchers thus obtain self-supporting silica films that have nanometre-sized channels, high surface areas and a twisting, rod-like substructure. The films exhibit iridescent colours similar to those of the original cellulose, and the colours can be tuned across the spectrum by modifying the

synthetic conditions. These films are promising new materials for applications as varied as chemical sensors, smart windows and separating molecules that differ only in terms of their 'handedness'.

Handedness, or chirality, is a geometrical property of molecules that can exist as isomers that are non-superimposable mirror images of each other. The property has considerable implications for the interactions of chiral molecules with other molecules. Just as a handshake feels right only if the hands are matched, chiral molecules interact best if their handedness matches.

Chirality is a widespread feature of biological



**Figure 1 | Capturing the chiral structure of liquid crystals in a solid film.** **a**, Cellulose nanoparticles with spindle-like shapes arrange themselves into chiral, helical liquid-crystal arrays. Only part of the helical structure is shown here. **b**, Shopsowitz *et al.*<sup>1</sup> report the optimal reaction conditions for preserving the helical arrays when cellulose nanoparticles are embedded in a silica matrix to produce composite materials. **c**, After eliminating the cellulose nanoparticles from the composites using heat, the authors obtained self-supporting films of silica that have a chiral porous structure.

systems, and is responsible for the specificity of many biological processes for particular substrates. Chiral substances also interact with polarized light, an effect that is used in liquid-crystal displays. These displays contain a layer of cigar-shaped molecules that can line up parallel to each other, like fish in a school swimming in the same direction, forming a nematic liquid crystal. But, in another state, they form a twisted, screw-like assembly. This twisting introduces handedness to the ensemble of molecules, so that polarized light travelling through the resulting chiral nematic liquid crystal is rotated.

Numerous applications can benefit from the incorporation of chirality into hard materials, including catalysis, molecular separation, chemical sensing and optics. The interactions of chiral materials with other chiral species are amplified if very large interfaces are used between them. Porous materials can provide the necessary large surface areas, particularly mesoporous materials — those with pore sizes of 2–50 nanometres. Such materials have attracted much interest because their pore morphology can be controlled using templates. Commonly used templates include surfactant-based micelles, which form various phases such as cylindrical or spherical arrays<sup>2,3</sup>. Inorganic structures can be assembled around micelles so that subsequent removal of the template yields porous materials.

Prior to Shopsowitz and colleagues' work<sup>1</sup>, a mesoporous material had been made<sup>4</sup> from silica using chiral surfactant molecules as templates. The product consisted of micrometre-long, twisted, rod-like structures of hexagonal cross-section and contained nanometre-sized channels that spiralled around the rods, reminiscent of fibres in a rope. Cheaper, naturally abundant templates, in particular cellulose nanocrystals<sup>5,6</sup>, have also been explored as alternatives to inducing chirality in mesoporous silica. These are easily obtained by acid treatment of bulk cellulose, which is present in wood pulp, cotton, green algae and other natural sources.

Cellulose nanocrystals have spindle-shaped structures, diameters of a few nanometres and screw symmetry. They are prone to lining up

to form nematic liquid crystals with helical structures. One would therefore expect nanocrystalline cellulose to be a suitable template for the nanocasting of chiral porous systems. Indeed, two early studies<sup>5,6</sup> hinted that chiral mesopores in silica could be formed using these cellulose derivatives as templates, but the authors were careful to note that the chiral domains may have been confined to small, localized regions. Obstacles to achieving long-range helical ordering have included the high sensitivity of the chiral phase of the cellulose templates to pH, concentration and temperature, and the tendency of the silica precursor used in the synthesis to disrupt the order of the nanocrystalline cellulose template<sup>7</sup>.

These obstacles have now been overcome by Shopsowitz and colleagues<sup>1</sup>. They report that careful optimization of the synthesis conditions, especially the pH of the reaction solution, permits the preparation of mesoporous silica films in which chiral ordering of the pores occurs throughout the film (Fig. 1). The films were self-supporting after the removal of the cellulose template, and had high surface areas of several hundred square metres per gram of material. The authors obtained scanning electron micrographs of film surfaces, from which the chiral nature of the material was evident from the arrangement of the twisted rods that make up the film (see Fig. 3d on page 423).

Shopsowitz *et al.* observed that the iridescent colours of the template were reproduced in their solid inorganic material, but varied with the fraction of silica precursor included in the synthesis mixture. The colours of the silica films could therefore be tuned across the visible spectrum to the near-infrared region. The authors carried out circular dichroism experiments on their films — that is, they illuminated the films with circularly polarized light in which the electric-field vector of the light beam traces a helix around the axis of the beam. They thus confirmed that the films' colours originate from the selective reflection of left-handed polarized light by the left-handed chiral nematic structure of the films. The colours disappeared when Shopsowitz *et al.* filled the mesopores with liquids that had a refractive index the same as

that of the silica walls, because this cancelled out the effects of the chiral pore geometry throughout the film.

This study opens up new opportunities for chiral solids because it demonstrates that long-range helical ordering of pores is achievable using inexpensive, renewable, chiral template materials, in a process that should be scalable. As the authors point out<sup>1</sup>, cellulose is already widely used to separate mixtures of chiral compounds. Mesoporous silica that reproduces the helical features of cellulose should therefore also benefit numerous applications that rely on chiral effects — perhaps more so than cellulose itself, because mesoporous silica is easily modified with other chemical groups<sup>8</sup>.

Shopsowitz and colleagues' silica films should be suitable templates for nanocasting replica films made of other materials<sup>9,10</sup>. Nanocrystalline cellulose templates could probably also be used directly to make films of different compositions, although the authors' research<sup>1</sup> shows that it will be necessary to find the optimal reaction conditions for each material used. What's more, they clearly demonstrate the effects of chiral pore structures on the optical properties of their films, which rely on feature sizes at the length scale of several hundred nanometres — the wavelength of visible light. Whether the chiral features in these materials will influence molecular interactions on a much smaller length scale remains to be seen. But now that a robust synthesis of chiral mesoporous materials has been developed, the necessary testing is feasible. ■

Andreas Stein is in the Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, USA. e-mail: a-stein@umn.edu

- Shopsowitz, K. E., Qi, H., Hamad, W. Y. & MacLachlan, M. J. *Nature* **468**, 422–425 (2010).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
- Zhao, D. *et al.* *Science* **279**, 548–552 (1998).
- Che, S. *et al.* *Nature* **429**, 281–284 (2004).
- Dujardin, E., Blaseby, M. & Mann, S. J. *Mater. Chem.* **13**, 696–699 (2003).
- Thomas, A. & Antonietti, M. *Adv. Funct. Mater.* **13**, 763–766 (2003).
- Pan, J., Hamad, W. & Straus, S. K. *Macromolecules* **43**, 3851–3858 (2010).
- Stein, A., Melde, B. J. & Schrodner, R. C. *Adv. Mater.* **12**, 1403–1419 (2000).
- Ryoo, R., Joo, S. H. & Jun, S. J. *Phys. Chem. B* **103**, 7743–7746 (1999).
- Stein, A., Wang, Z. & Fierke, M. A. *Adv. Mater.* **21**, 265–293 (2009).

#### CORRECTION

In the News & Views article 'Biological physics: Filaments band together' by Jean-François Joanny and Sriram Ramaswamy (*Nature* **467**, 33–34; 2010), reference 12 (Chaté, H. *et al.* *Phys. Rev. E* **77**, 046113; 2008), not reference 11, should have been cited as the source of the image in Figure 1a.



# The moment of truth for WIMP dark matter

Gianfranco Bertone<sup>1,2</sup>

**We know that dark matter constitutes 85 per cent of all the matter in the Universe, but we do not know of what it is made. Amongst the many dark matter candidates proposed, WIMPs (weakly interacting massive particles) occupy a special place, because they arise naturally from new theories that seek to extend the standard model of particle physics. With the advent of the Large Hadron Collider at CERN, and a new generation of astroparticle experiments, the moment of truth has come for WIMPs: either we will discover them in the next five to ten years, or we will witness their inevitable decline.**

The foundations of the modern dark matter problem<sup>1</sup> were laid in the 1970s and 1980s, after decades of slow accumulation of evidence<sup>2</sup>. It was noticed in the 1930s that the Coma cluster seemed to contain much more mass than could be inferred from visible galaxies<sup>3</sup>, and a few years later, it became clear that the Andromeda galaxy rotates very fast at large radii, as if most of its mass lay in its outer regions<sup>4</sup>. Several other pieces of evidence provided further support to the dark matter hypothesis, including the so-called timing argument<sup>5,6</sup>, until in the 1970s rotation curves were extended to larger radii and to many other spiral galaxies, proving the presence of large amounts of mass on scales much larger than the size of galactic disks<sup>7,8</sup>. Although it is in principle possible to explain these observations in terms of new theories of gravity<sup>9</sup> (after all, we only have gravitational evidence for dark matter), lensing observations of galaxy clusters provide a formidable challenge to these theories<sup>10,11</sup>.

Today, we have entered the era of precision cosmology: we can determine the abundance of dark matter in the Universe with exquisite accuracy<sup>12</sup>; we have a much better understanding of how dark matter is distributed in structures that range from dwarf galaxies to clusters of galaxies, thanks to both high-resolution numerical simulations made possible by modern supercomputers<sup>13</sup> and lensing observations<sup>14</sup>; and we even have a rather precise idea of how the Milky Way formed, and of the local abundance of dark matter<sup>15,16</sup>. More importantly, we know today that dark matter cannot be made of ordinary matter, so new particles must exist<sup>17</sup>, unless we are completely misled by a wide array of astrophysical and cosmological observations.

Particle physicists have proposed literally tens of possible dark matter candidates. Axions, for instance, are hypothetical particles whose existence was postulated to solve the so-called strong CP problem in quantum chromodynamics, and they are known to be very well motivated dark matter candidates<sup>18,19</sup>. Other well-known candidates are sterile neutrinos, which interact only gravitationally with ordinary matter, apart from a small mixing with the familiar neutrinos of the standard model<sup>20,21</sup>. A wide array of other possibilities have been discussed in the literature, and they are currently being searched for with a variety of experimental strategies<sup>22,23</sup>.

The most studied class of candidates, however, is that of WIMPs, which have the virtue of naturally achieving the correct relic abundance (see Box 1) in the early Universe. The reason they became so popular is that WIMP candidates arise naturally from theories that seek to extend the standard model of particle physics, and to embed it in a more fundamental theory. In particular, it was noticed back in 1983 that one of the most promising extensions of the standard model, supersymmetry, provides an excellent dark matter candidate: the neutralino<sup>24–27</sup>. This particle

fulfils all the properties of the good dark matter candidate, and it has become over the years a prototypical example of a WIMP. Its mass can range from about 50 GeV (I adopt here units of  $c = 1$ , unless otherwise specified) to a few TeV, and its interaction cross-section with ordinary matter and with itself are such that it can account for all the dark matter in the Universe while still remaining consistent with all known experiments.

If dark matter is made of WIMPs, we should be able to detect it. We could in principle observe the interaction of dark matter particles with nuclei in underground detectors, as proposed back in 1985<sup>28</sup>, or we may detect the products of annihilation or decay of these particles, as first discussed almost three decades ago<sup>23,29,30</sup>. Although all the search strategies so far devised have failed to provide incontrovertible evidence for dark matter particles, today a new generation of particle astrophysics experiments is about to start, or has already started, taking data. Furthermore,

## BOX 1

## WIMPs

In the simplest WIMP models, dark matter particles are kept in thermal and chemical equilibrium in the early Universe with all other particles, by virtue of their self-annihilation into particles of the standard model and vice versa. Their density rapidly decreases as the Universe expands, until it becomes so low that WIMPs cannot self-annihilate any more and they freeze-out from equilibrium, that is, their co-moving number density remains fixed. Under some simplifying assumptions<sup>23</sup>, the relic abundance of WIMPs in the Universe  $\Omega_\chi h^2$  (that is, the number density of WIMPs in the local Universe in units of the critical density (see ref. 12 for further details)) can be simply expressed in terms of the self-annihilation cross-section,  $\sigma v$ :

$$\Omega_\chi h^2 \approx \frac{3 \times 10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\sigma v} \quad (1)$$

where  $h$  is the Hubble parameter, which encodes the expansion rate of the Universe, in units of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . As the measured value of  $\Omega_\chi h^2$  is around 0.1 (ref. 12), the self-annihilation cross-section required in order to achieve the appropriate relic density is  $\sigma v \approx 3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ , a cross-section typical of weak interactions in the standard model, hence the name WIMPs. Although in this simplified calculation the relic density does not depend strongly on the mass of the dark matter particle,  $m_\chi$ , the maximum and minimum annihilation cross-sections of the most common candidates do depend on it, therefore constraining the values of  $m_\chi$  (in GeV) to the range  $10 \lesssim m_\chi \lesssim 10^5$  (ref. 23).

<sup>1</sup>Institut d'Astrophysique de Paris, UMR 7095-CNRS, Université Pierre et Marie Curie, 98 bis Boulevard Arago, 75014 Paris, France. <sup>2</sup>Institute für Theoretische Physik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

the Large Hadron Collider (LHC) at CERN has recently started operations, and it is expected to find, or to severely constrain, the most studied extensions of the standard model, including supersymmetry.

I argue here that the moment of truth has therefore come for WIMP dark matter, for we will either discover them at the LHC and in particle astrophysics experiments in the next 5 to 10 years, or the case for WIMPs will become weak, and we will witness their inevitable decline.

## Indirect detection

Indirect detection consists of the search for the annihilation or decay products of dark matter particles, such as photons, antimatter and neutrinos. WIMPs in fact are expected to self-annihilate efficiently in regions where they accumulate, such as the centre of galactic haloes or substructures such as dwarf galaxies, as the annihilation rate depends on the square of the number density. Once they annihilate, they produce secondary particles, such as quarks and gauge bosons, which subsequently fragment and decay in the aforementioned final states (photons and so on). The typical energy of these final states is about a tenth of the dark matter particle mass, so we can search indirectly for dark matter by looking for an excess of photons, antimatter or neutrinos in astrophysical data at energies between 1 GeV and 10 TeV (Box 1).

Obtaining convincing evidence for dark matter from astrophysical observations has proved a very difficult task. It is in fact easy to model almost any excess in the measured energy spectrum of photons or antimatter, at any energy, in terms of dark matter particles with suitable properties. One simply has to follow three steps: (1) adjust the normalization of the flux by changing the distribution of dark matter particles and their annihilation cross-section; (2) choose a dark matter mass that provides the correct energy scale; and (3) fit the spectral features by choosing an appropriate annihilation channel and, in the case of antimatter, by tuning the propagation parameters. In practice, there is enough freedom to fit almost any astrophysical observation; and in fact, features in the data of many experiments of the past 5–10 years have been tentatively interpreted in terms of different dark matter candidates, sometimes even at the cost of making unrealistic assumptions about the nature and distribution of dark matter.

The most recent example is the rise in the energy spectrum of the positron ratio—that is, the number of positrons divided by the sum of the numbers of positrons and electrons—measured by the PAMELA satellite above 10 GeV (ref. 31). The standard WIMP model (that is, a particle with a mass in the range  $10^2$ – $10^3$  GeV, and a thermal cross-section  $\sigma v \approx 10^{-26} \text{ cm}^3 \text{ s}^{-1}$ , where  $\sigma$  is the self-annihilation cross-section and  $v$  is the relative velocity of WIMPs) can hardly account for this feature, so new *ad hoc* candidates have been proposed: particles with a very large annihilation cross-section (high enough to match the normalization of the positron ratio, but not too much, in order to avoid cosmological constraints<sup>32,33</sup>), annihilating only to leptons (to evade anti-proton constraints<sup>34</sup>), and with a density profile shallower than that suggested by numerical simulations (to evade  $\gamma$ -ray constraints from the Galactic Centre<sup>35</sup>). There is therefore a possible combination of parameters that can be made compatible with all observations, but this is certainly not enough to claim discovery of dark matter, for there are less-exotic astrophysical sources that can account for the same feature without invoking new particles with *ad hoc* properties.

Fortunately, there are actually a number of astrophysical observations that might lead to convincing evidence, in the sense that they could be explained only in terms of dark matter, while being incompatible with a standard astrophysical interpretation. A typical example of such ‘smoking gun’ evidence would be the observation of a high-energy  $\gamma$ -ray line, which would point directly to the existence of new particles annihilating directly to photons. In fact, if WIMPs do not produce photons through the fragmentation and decay of secondary particles, but do so directly, the photons produced in the annihilation will be mono-energetic, thus producing a line in the  $\gamma$ -ray spectrum at an energy equal to the mass of the dark matter particle. The Fermi LAT satellite however did not observe such lines<sup>36</sup>; and it excluded cross-sections for annihilation to

photons that are larger than the thermal cross-section. We can expect an improvement in sensitivity of the Fermi LAT to  $\gamma$ -ray lines of one order of magnitude at most over the next decade (at least in the energy range where the sensitivity is limited by statistics, and not by the background, in which case the sensitivity scales with the square root of time).

Another very clean signature of dark matter annihilations would be the observation of high-energy neutrinos from the centre of the Sun<sup>30</sup>. Solar neutrinos produced in nuclear reactions have energies in the MeV range, so the observation of  $10^2$ – $10^4$  GeV neutrinos would require an explanation in terms of new physics, and the well studied process of capture and annihilation of dark matter particles in the Sun would provide it. The problem is that the neutrino telescope IceCube, currently under construction at the South Pole, so far has not found any evidence for an excess of neutrinos from the Sun. Over the next 5 years, the experiment will improve its sensitivity by a factor of  $\sim 5$ , and extend the threshold down to 50 GeV, with the construction of a more densely instrumented portion of detector, called DeepCore<sup>37</sup>. Even with these technical improvements and longer exposure, most of the supersymmetry parameter space will remain inaccessible, and the same holds true for the so-called Kaluza–Klein dark matter in theories with universal extra dimensions, where all particles and fields can propagate in new dimensions beyond the  $3+1$  we are familiar with.

Other strategies may provide useful hints. One is the multi-wavelength approach, which consists of the combined analysis of astrophysical spectra at different wavelengths. For example, dark matter annihilations produce  $\gamma$ -rays, but they also produce secondary electrons that could give rise to synchrotron and inverse Compton emission<sup>38</sup>. Another is the study of the angular power spectrum of  $\gamma$ -ray anisotropies<sup>39</sup>; this may allow the identification of a dark matter contribution to the diffuse  $\gamma$ -ray background. But even in the case of detection, it would probably require a long time before these observations are considered proof of the existence of dark matter, because one would have to exclude an astrophysical origin of the signal. Fortunately, although indirect searches may appear to be not particularly suited to the provision of incontrovertible evidence for dark matter, they have the big advantage of not requiring dedicated experiments; also, some theoretical models are indeed within the reach of current and upcoming experiments in the next 5–10 years. In the absence of these (admittedly optimistic) ‘smoking gun’ observations, a convincing case for dark matter can be made only if there are successful searches at accelerators or if direct detection experiments prove successful—in which case indirect searches may still provide useful information on the distribution of dark matter.

## Direct detection

The field of direct detection appears perhaps in a better shape, given the prospects of increasing the size, and therefore the sensitivity, of current experiments by at least two orders of magnitude within 5–10 years. The idea is to detect the recoil energy of nuclei struck by dark matter particles travelling through a detector, through the measurement of the light, the charge or the phonons produced in the target material by the scattering event. The progress made in this field is rather spectacular: the sensitivity of direct detection experiments (in terms of detectable upper limits) has gone down by more than three orders of magnitude in the past 20 years (ref. 40). Despite the extraordinary technological progress, however, dark matter has not yet been identified.

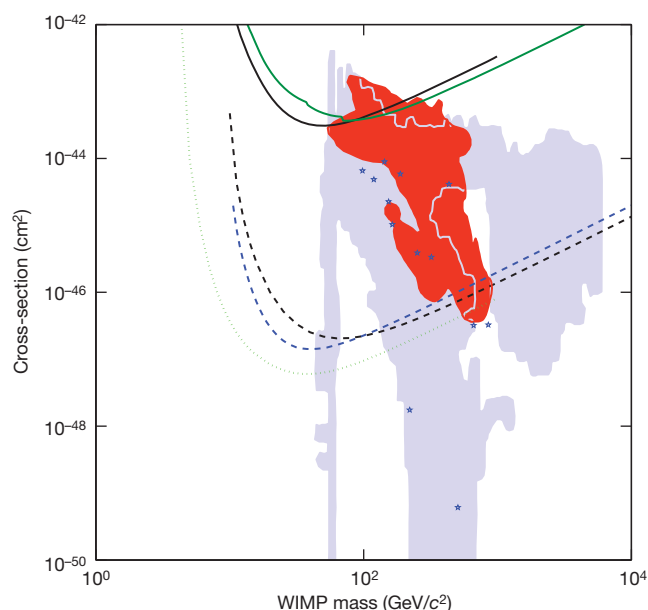
Something similar to the case of indirect detection actually happens for direct searches: there are intriguing signals that might be interpreted in terms of dark matter, but the case for WIMPs is simply not strong enough to convince the community. The best-known and most discussed example is the DAMA/LIBRA experiment, which reported the detection of a yearly modulation in the measured event rate, compatible with what is expected in common dark matter models, where the rate is modulated by the Earth’s revolution around the Sun<sup>41</sup>. However, its interpretation in terms of the elastic scattering of a WIMP ( $\chi$ , with a mass around 10–100 GeV) off a proton ( $p$ ) with a scalar, or spin-independent (SI), cross-section of  $\sigma_{\chi-p}^{\text{SI}} \approx 10^{-3}$ – $10^{-5}$  pb, has been challenged by



other experiments<sup>42</sup>. The CoGeNT collaboration also recently reported an excess of low-energy events which could be explained in terms of a very light WIMP<sup>43</sup>, but this interpretation is in tension with the first XENON100 results<sup>44</sup> (the debate on the DAMA/LIBRA and CoGeNT results is still open<sup>45</sup>). Finally, the CDMS II collaboration recently announced the detection of two events compatible with a WIMP signal; but this is still far from a discovery, as the expected background was 0.8 events<sup>46</sup>.

What can we expect from future experiments, and how do we get convincing evidence? The first step would be the detection of a rate of events significantly larger (that is, 5 standard deviations) than the expected background, as determined before the unblinding of the data. One could then try to assess the WIMP mass and scattering cross-sections compatible with the measured rate. Given the small number of events of the first detection, the reconstruction procedure would probably be rather poor—in the sense that the data will not set stringent constraints on the WIMP mass and cross-section—unless the mass of the dark matter particle is very small<sup>47</sup>.

At that point, in order to validate the dark matter interpretation, it would be crucial to add an independent piece of evidence that would provide strong support to the first discovery claim. This could be a discovery in an accelerator, as we shall see in the next section, or a direct detection in a different experiment—the latter would be better if it involved a different target material. This independent discovery would allow a much more precise determination of the WIMP mass<sup>48</sup>, and be an effective way to discriminate among WIMP candidates<sup>49</sup>.

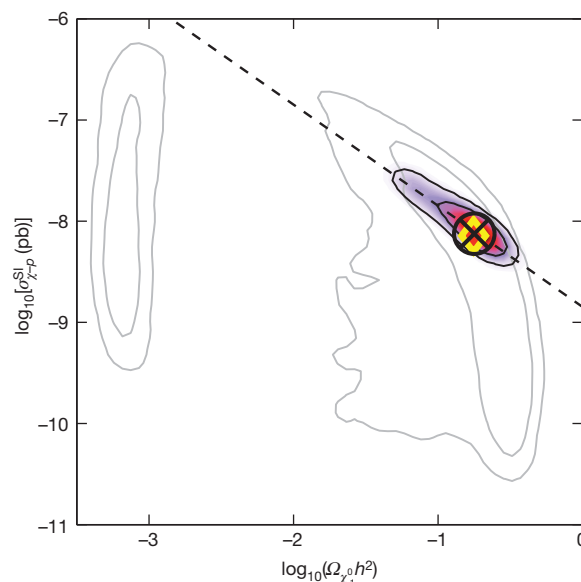


**Figure 1 | Status of direct dark matter searches.** The figure shows the status of direct dark matter searches (lines and coloured areas) on a plot of the scalar WIMP–proton scattering cross-section versus WIMP mass. Current experiments have excluded models above the solid lines (CDMS II, green line; Xenon100, black line). The reach of several upcoming experiments is shown by the dashed lines (black, SuperCDMS Phase C with 1 ton of germanium; blue, LUX with 3 tons of liquid xenon; and green, Xenon1T with 1 ton of xenon). Also shown for comparison are the predictions for different theoretical models. Stars correspond to benchmark models (corresponding to typical regions in the theoretical parameter space) in a constrained supersymmetric set-up with only four free parameters (see text for further details, in particular Box 2)<sup>50</sup>. The most probable region of the parameter space for this theoretical set-up can be determined with a Monte Carlo Markov Chain procedure<sup>51</sup>, and it is shown here in red. Finally, we also show (blue area) the result of a scan of the parameter space performed in a less constrained supersymmetric set-up with seven free parameters specified at low energy<sup>52</sup>, to stress the fact that a more rich phenomenology is in general allowed by supersymmetric theories. The plot has been made with DMTools (<http://dmttools.brown.edu>).

In Fig. 1 I summarize the current situation of direct dark matter searches. The figure shows the sensitivity of current and upcoming experiments, compared with theoretical predictions, in the WIMP–proton cross-section. The cross-section shown in Figs 1 and 2 corresponds to a scalar (or spin-independent) coupling. WIMPs can also interact with the spin of the nucleon, with an axial (or spin-dependent) coupling. The theoretical predictions depend on the specific model considered.

The small stars in Fig. 1, for instance, correspond to a set of benchmark models in a supersymmetric theoretical set-up called minimal supergravity (mSUGRA)<sup>50</sup> (see Box 2 for further details on theoretical models). Another very similar supersymmetric set-up is the constrained minimal supersymmetric model; the red area in Fig. 1 shows the most probable region of the parameter space of this theoretical set-up, as determined with a Monte Carlo Markov Chain procedure<sup>51</sup>. It is worth noting that these predictions have been made in the framework of a constrained version of a more general class of supersymmetric models, that allow in general a much more rich phenomenology: the blue area in Fig. 1 shows the result of a similar Markov Chain scan in the framework of a more general supersymmetric model with seven free parameters<sup>52</sup> (see Box 2).

As we can see in Fig. 1, a large portion of the parameter space where theoretical models lie will be probed by ton-scale experiments that are expected to start within 5–10 years. This is good news, as for this set of parameters we can perform the program described above. But we have to consider the possibility that supersymmetry, or in general the dark matter particle, is outside the reach of ton-scale experiments. In this case, the question will arise of whether one should continue searching, and build even bigger and more expensive detectors, or simply stop, and



**Figure 2 | Complementarity between accelerator and direct detection searches.** We show a reconstruction of the properties of the dark matter particle on a plot of the scattering cross-section versus relic density  $\Omega_{\chi_1^0}$ , starting from the benchmark point indicated by the yellow/red diamond (see text for further details). This model is within the reach of the LHC, so we can simulate the set of measurements that should become available with, say, 300 fb<sup>-1</sup> of data, corresponding roughly to the data that will be accumulated by 2016, if the experiment runs according to plan. The result of the reconstruction procedure based on LHC data only, as performed in a supersymmetric set-up with 24 free parameters (see Box 2), is shown by the light grey contours, which exhibit a double peak structure, with a very broad peak around the true value. Fortunately, in the case of direct detection with a ton-scale experiment, the reconstruction procedure becomes much more precise, as shown by the coloured areas within the black contours, as this type of experiment breaks the degeneracy in the parameter space along the dashed line. In this case, the best fit point, shown by the encircled black cross, practically lies on top of the true value<sup>57</sup>.

## BOX 2

## Beyond the standard model

The standard model of particle physics is viewed by many as an effective field theory valid for energies up to the TeV scale, rather than a truly fundamental theory. This belief is not based on discrepancies with experimental results, but on (very strong) theoretical arguments. Among them, the so-called hierarchy problem is perhaps the most prominent: in order to stabilize the mass of the Higgs against quadratically divergent radiative corrections without an unacceptable amount of fine-tuning (an adjustment of 32 orders of magnitude, for the standard model to be valid up to the Planck scale), the scale of new physics must be  $\mathcal{O}(1)$  TeV, that is, within the reach of the LHC (see, for example, ref. 54). As reasonable and aesthetically appealing as it is, this is not a rigorous mathematical argument, and the actual scale of new physics could be in principle even higher, depending on the amount of fine-tuning one is willing to tolerate.

Among the proposed extensions of the standard model, supersymmetry is undoubtedly the most studied, and probably one of the best motivated. The so called minimal supersymmetric standard model has however about 120 free parameters, and although not all of them are relevant for the calculation of the properties of dark matter candidates, some assumptions must be made about the structure of the theory in order to reduce the number of free parameters, and make quantitative predictions for the mass and couplings of supersymmetric particles. In this Review I refer to some of the most popular supersymmetric models:

- The constrained minimal supersymmetric model (cMSSM) and the minimal supergravity (mSUGRA) model are supersymmetric theories with four free parameters. They differ in some small technical details, but both are often used to make predictions for dark matter searches, because despite the very strong theoretical assumptions made to reduce the number of free parameters (that is, universality of masses and couplings at the grand unification scale), they capture the main aspects of the phenomenology of supersymmetric theories.

- The phenomenological supersymmetric model is a phenomenological model that is specifically tailored to the study of dark matter. There are different versions of the model, one of the most popular being a seven free parameters theory, where all parameters are specified at low energy, as implemented in the popular DarkSUSY code<sup>58</sup>. A less constrained version of this model has 24 free parameters, and it is the one adopted here to discuss the complementarity of direct and accelerator searches (see in particular Fig. 2).

focus on something different. The answer will probably depend on what is found in accelerators, as we shall see in the next section. But it is worth recalling that coherent interactions of neutrinos provide an irreducible background for these searches, thereby limiting the capability to probe very low scattering cross-sections<sup>53</sup>.

## Accelerators

The detection strategy that appears perhaps most promising today is the search for new physics in accelerators. There are in fact high expectations for the LHC, which has recently started operations at CERN. The current plan is to run it at a centre of mass energy of 7 TeV until the end of 2011, and then, after an upgrading procedure, at 14 TeV. The existence of new particles at the TeV scale can be tested—most theorists believe that signs of new physics should appear at these energies.

Among the proposed extensions of the standard model, supersymmetry is undoubtedly the most studied, and probably one of the best motivated. Not only does it solve the hierarchy problem (Box 2) in a natural way, but it also provides a perfect dark matter candidate (more than one, in fact<sup>54</sup>). There are large portions of the supersymmetric parameter space within the reach of the LHC, and there are good chances of discovering it at this accelerator within 5–10 years (ref. 55).

There is no need to stress the impact that a detection of new physics would have on our description of the Universe. I limit myself here to discussing the consequences for dark matter searches. In particular, a natural question to ask is: how do we understand whether newly discovered particles have something to do with the dark matter in the Universe? From accelerator measurements, we can infer the existence of a particle that is stable over the timescale it takes for it to escape the detector, that is, less than 1  $\mu$ s. But we cannot prove that it is stable over cosmological timescales, nor we can assess its relic density in the absence of a theoretical framework in which to perform the calculation of the cosmological evolution of its density.

Even if supersymmetry is discovered, and the mass spectrum of new particles is determined with good accuracy, reconstructing the relic density of the neutralino will be challenging<sup>56</sup>, unless the analysis is performed in a low-dimensional parameter space (for example, mSUGRA<sup>55</sup>). In the ‘dream’ scenario where new physics is discovered at the LHC, fortunately, particle astrophysics experiments can provide complementary information on the nature of dark matter. In fact, direct searches provide an effective way to reduce degeneracies in the parameter space of new theories, when reasonable assumptions are made about the distribution of dark matter particles in the Milky Way. We show in Fig. 2 an example of a recent study in the framework of a 24-parameter supersymmetric set-up (Box 2), where the simulated response of the LHC and of 1-ton experiments to a given benchmark model was used to reconstruct the relic density of dark matter. The study showed that a convincing identification of dark matter particles is possible with a combination of LHC and direct detection data<sup>57</sup>.

## The future

The other possibility is of course that new physics is not found at the LHC within 5–10 years. For the reasons I have discussed above, null searches at the LHC would push the scale of new physics into more and more unnatural territory (that is, to high levels of fine-tuning). Although null searches would not rule out supersymmetry and many other new theories, they would cast doubt on the very existence of new physics at any scale, especially if the Higgs boson is found, completing the standard model.

For WIMP dark matter studies, the consequences would be dramatic. In the absence of new colliders—for which we would have to wait at least 20 years—the only remaining hope would be to obtain ‘smoking gun’ evidence from direct or indirect detection. But indirect searches are complicated, as we have seen, and even assuming that one can make a strong case for supersymmetry (for example) at a higher scale, they are actually much more difficult for high-mass dark matter particles. This is because the annihilation spectra scale with the inverse of the mass squared, and also because in general the detection of photons and anti-matter is difficult at energies above tens of TeV. As for direct detection, in the absence of any trace of new physics at the LHC, it will be probably difficult to motivate the construction of experiments beyond the ton scale. In the absence of any signal, we would be left with the ‘nightmare’ dark matter scenario of null searches at the LHC, and no direct or indirect detections. Such circumstances would probably mark the decline of WIMPs in favour of alternative explanations, such as axions or alternative theories of gravity, provided that they can be reconciled with lensing observations.

Let us stay optimistic, though. The plans to detect dark matter in the near future have been laid out carefully, and they deserve to be carried out with the utmost care. A discovery would mark the start of a new era of physics, and it would represent the best reward for decades of painstaking searches.

1. Bertone, G. (ed.) *Particle Dark Matter: Observations, Models and Searches* (Cambridge Univ. Press, 2010).  
**Contains an updated discussion of the various aspects of the dark matter problem in astrophysics, cosmology and particle physics.**
2. Einasto, J. Dark matter. In *UNESCO EOLSS Encyclopedia*; preprint at (<http://arXiv.org/abs/0901.0632>) (2009).



3. Zwicky, F. Spectral displacement of extra galactic nebulae. *Helv. Phys. Acta* **6**, 110–127 (1933).
4. Babcock, H. W. The rotation of the Andromeda Nebula. *Lick Obs. Bull.* **19**, 41–51 (1939).
5. Kahn, F. D. & Woltjer, L. Intergalactic matter and the galaxy. *Astrophys. J.* **130**, 705–717 (1959).
6. Einasto, J. & Lynden-Bell, D. On the mass of the Local Group and the motion of its barycentre. *Mon. Not. R. Astron. Soc.* **199**, 67–80 (1982).
7. Bosma, A. *The Distribution and Kinematics of Neutral Hydrogen in Spiral Galaxies of Various Morphological Types*. PhD thesis, Groningen Univ. (1978).
8. Rubin, V. C., Ford, W. K. J. & Thonnard, N. Rotational properties of 21 SC galaxies with a large range of luminosities and radii. *Astrophys. J.* **238**, 471–487 (1980).
9. Milgrom, M. & Bekenstein, J. in *Dark Matter in the Universe* (eds Kormendy, J. & Knapp, G. R.) 319–330 (IAU Symp. No. 117, 1987).
10. Clowe, D. *et al.* A direct empirical proof of the existence of dark matter. *Astrophys. J.* **648**, L109–L113 (2006).
11. Angus, G. W., Shan, H., Zhao, H. & Famaey, B. On the law of gravity, the mass of neutrinos and the proof of dark matter. *Astrophys. J.* **654**, L13–L16 (2007).
12. Komatsu, E. *et al.* Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation. Preprint at (<http://arXiv.org/abs/1001.4538>) (2010).
- Contains a state of the art determination of cosmological parameters, and shows that dark matter is a necessary ingredient of the standard cosmological model.**
13. Diemand, J. & Moore, B. in *Particle Dark Matter: Observations, Models and Searches* (ed. Bertone, G.) 14–37 (Cambridge Univ. Press, 2010).
14. Mellier, Y. in *Particle Dark Matter: Observations, Models and Searches* (ed. Bertone, G.) 56–82 (Cambridge Univ. Press, 2010).
15. Catena, R. & Ullio, P. A novel determination of the local dark matter density. *J. Cosmol. Astropart. Phys.* **08**, 004 (2010).
16. Pato, M., Agertz, O., Bertone, G., Moore, B. & Teyssier, R. Systematic uncertainties in the determination of the local dark matter density. *Phys. Rev. D* **82**, 023531 (2010).
17. Taoso, M., Bertone, G. & Masiero A.. Dark matter candidates: a ten-point test. *J. Cosmol. Astropart. Phys.* **03**, 022 (2008).
18. Sikivie, P. in *Particle Dark Matter: Observations, Models and Searches* (ed. Bertone, G.) 204–227 (Cambridge Univ. Press, 2010).
19. Visinelli, L. & Gondolo, P. Dark matter axions revisited. *Phys. Rev. D* **80**, 035024 (2009).
20. Shaposhnikov, M. in *Particle Dark Matter: Observations, Models and Searches* (ed. Bertone, G.) 228–248 (Cambridge Univ. Press, 2010).
21. Boyarsky, A., Ruchayskiy, O. & Shaposhnikov, M. The role of sterile neutrinos in cosmology and astrophysics. *Annu. Rev. Nucl. Part. Sci.* **59**, 191–214 (2009).
22. Bergström, L. Non-baryonic dark matter: observational evidence and detection methods. *Rep. Prog. Phys.* **63**, 793–841 (2000).
23. Bertone, G., Hooper, D. & Silk, J. Particle dark matter: evidence, candidates and constraints. *Phys. Rep.* **405**, 279–390 (2005).
24. Goldberg, H. Constraint on the photino mass from cosmology. *Phys. Rev. Lett.* **50**, 1419–1422 (1983); erratum. *Phys. Rev. Lett.* **103**, 099905 (2009).
25. Ellis, J. R., Hagelin, J. S., Nanopoulos, D. V., Olive, K. A. & Srednicki, M. Supersymmetric relics from the big bang. *Nucl. Phys. B* **238**, 453–476 (1984).
26. Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. Formation of galaxies and large-scale structure with cold dark matter. *Nature* **311**, 517–525 (1984).
27. Jungman, G., Kamionkowski, M. & Griest, K. Supersymmetric dark matter. *Phys. Rep.* **267**, 195–373 (1996).
28. Goodman, M. W. & Witten, E. Detectability of certain dark-matter candidates. *Phys. Rev. D* **31**, 3059–3063 (1985).
29. Silk, J. & Srednicki, M. Cosmic-ray antiprotons as a probe of a photino-dominated universe. *Phys. Rev. Lett.* **53**, 624–627 (1984).
30. Silk, J., Olive, K. A. & Srednicki, M. The photino, the sun, and high-energy neutrinos. *Phys. Rev. Lett.* **55**, 257–259 (1985).
31. Adriani, O. *et al.* An anomalous positron abundance in cosmic rays with energies 1.5–100 GeV. *Nature* **458**, 607–609 (2009).
- Discusses the discovery of an anomalous abundance of positrons with the PAMELA anti-matter satellite, tentatively interpreted in terms of the annihilation of dark matter particles.**
32. Galli, S., Iocco, F., Bertone, G. & Melchiorri, A. CMB constraints on Dark Matter models with large annihilation cross-section. *Phys. Rev. D* **80**, 023505 (2009).
33. Slatyer, T. R., Padmanabhan, N. & Finkbeiner, D. P. CMB constraints on WIMP annihilation: energy absorption during the recombination epoch. *Phys. Rev. D* **80**, 043526 (2009).
34. Cirelli, M., Kadastik, M., Raidal, M. & Strumia, A. Model-independent implications of the  $e^+e^-$  cosmic ray spectra on properties of Dark Matter. *Nucl. Phys. B* **813**, 1–21 (2009).
35. Bertone, G., Cirelli, M., Strumia, A. & Taoso, M. Gamma-ray and radio tests of the  $e^+e^-$  excess from DM annihilations. *J. Cosmol. Astropart. Phys.* **03**, 009 (2009).
36. Abdo, A. A. *et al.* Fermi LAT search for photon lines from 30 to 200 GeV and dark matter implications. *Phys. Rev. Lett.* **104**, 091302 (2010).
37. Halzen, F. & Hooper, D. The indirect search for dark matter with IceCube. *New J. Phys.* **11**, 105019 (2009).
38. Profumo, S. & Ullio, P. in *Particle Dark Matter: Observations, Models and Searches* (ed. Bertone, G.) 547–564 (Cambridge Univ. Press, 2010).
39. Ando, S. & Komatsu, E. Anisotropy of the cosmic gamma-ray background from dark matter. *Phys. Rev. D* **73**, 023521 (2006).
40. Gaitskill, R. Direct detection of dark matter. *Annu. Rev. Nucl. Part. Sci.* **54**, 315–359 (2004).
41. Bernabei, R. *et al.* First results from DAMA/LIBRA and the combined results with DAMA/Nal. *Eur. Phys. J. C* **56**, 333–355 (2008).
42. Fornengo, N. in *Particle Dark Matter: Observations, Models and Searches* (ed. Bertone, G.) 383–391 (Cambridge Univ. Press, 2010).
43. Aalseth, C. E. *et al.* Results from a search for light-mass dark matter with a P-type point. Preprint at (<http://arXiv.org/abs/1002.4703>) (2010).
44. Aprile, E. *et al.* First dark matter results from the XENON100 experiment. Preprint at (<http://arXiv.org/abs/1005.0380>) (2010).
45. Savage, C., Gelmini, G., Gondolo, P. & Freese, K. XENON10/100 dark matter constraints in comparison with CoGeNT and DAMA: examining the  $L_{eff}$  dependence. Preprint at (<http://arXiv.org/abs/1006.0972>) (2010).
46. Ahmed, Z. *et al.* Results from the final exposure of the CDMS II experiment. Preprint at (<http://arXiv.org/abs/0912.3592>) (2009).
47. Green, A. M. Determining the WIMP mass from a single direct detection experiment, a more detailed study. *J. Cosmol. Astropart. Phys.* **07**, 005 (2008).
48. Drees, M. & Shan, C. L. Model-independent determination of the WIMP mass from direct dark matter. *J. Cosmol. Astropart. Phys.* **06**, 012 (2008).
49. Bertone, G., Cerdeño, D. G., Collar, J. I. & Odom, B. C. WIMP identification through a combined measurement of axial and scalar couplings. *Phys. Rev. Lett.* **99**, 151301 (2007).
50. Battaglia, M. *et al.* Updated post-WMAP benchmarks for supersymmetry. *Eur. Phys. J. C* **33**, 273–296 (2004).
51. Trotta, R., Feroz, F., Hobson, M. P., Roszkowski, L. & Ruiz de Austri, R. The impact of priors and observables on parameter inferences in the constrained MSSM. *J. High Energy Phys.* **12**, 024 (2008).
52. Baltz, E. A. & Gondolo, P. Markov chain Monte Carlo exploration of minimal supergravity with implications for dark matter. *J. High Energy Phys.* **0410**, 052 (2004).
53. Strigari, L. E. Neutrino coherent scattering rates at direct dark matter detectors. *New J. Phys.* **11**, 105011 (2009).
54. Ellis, J. & Olive, K. A. in *Particle Dark Matter: Observations, Models and Searches* (ed. Bertone, G.) 142–162 (Cambridge Univ. Press, 2010).
55. Nath, P. *et al.* The hunt for new physics at the Large Hadron Collider. *Nucl. Phys. Proc., Suppl.* **200–202**, 185 (2010).
56. Baltz, E. A., Battaglia, M., Peskin, M. E. & Wizansky, T. Determination of dark matter properties at high-energy colliders. *Phys. Rev. D* **74**, 103521 (2006).
57. Bertone, G., Cerdeno, D. G., Fornasa, M., de Austri, R. R. & Trotta, R. Identification of dark matter particles with LHC and direct detection data. *Phys. Rev. D* (in the press); preprint at (<http://arXiv.org/abs/1005.4280>) (2010).
- Shows how to combine in a consistent way data from direct and accelerator searches, and highlights the complementarity of these two search strategies.**
58. Gondolo P. *et al.* DarkSUSY: Computing supersymmetric dark matter properties numerically. *J. Cosmol. Astropart. Phys.* **07**, 008 (2004).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to the author ([bertone@iap.fr](mailto:bertone@iap.fr)).

# Support for a synaptic chain model of neuronal sequence generation

Michael A. Long<sup>1†</sup>, Dezhe Z. Jin<sup>2</sup> & Michale S. Fee<sup>1</sup>

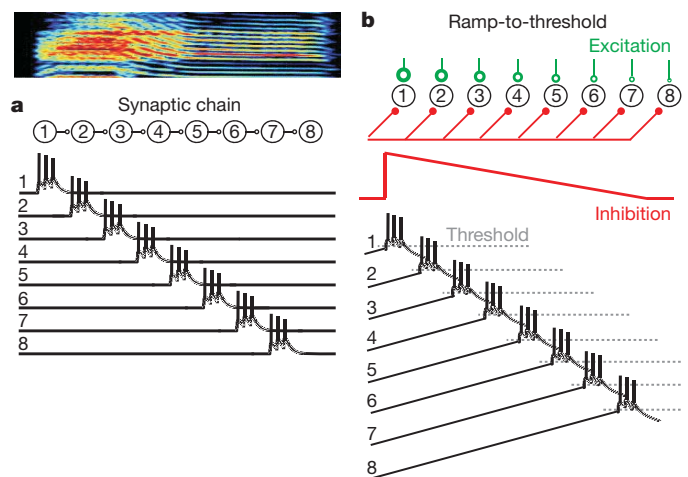
**In songbirds, the remarkable temporal precision of song is generated by a sparse sequence of bursts in the premotor nucleus HVC. To distinguish between two possible classes of models of neural sequence generation, we carried out intracellular recordings of HVC neurons in singing zebra finches (*Taeniopygia guttata*). We found that the subthreshold membrane potential is characterized by a large, rapid depolarization 5–10 ms before burst onset, consistent with a synaptically connected chain of neurons in HVC. We found no evidence for the slow membrane potential modulation predicted by models in which burst timing is controlled by subthreshold dynamics. Furthermore, bursts ride on an underlying depolarization of ~10-ms duration, probably the result of a regenerative calcium spike within HVC neurons that could facilitate the propagation of activity through a chain network with high temporal precision. Our results provide insight into the fundamental mechanisms by which neural circuits can generate complex sequential behaviours.**

Complex behaviours are made possible by the ability of the brain to step through well defined sequences of neural states<sup>1</sup>. Brain processes capable of generating intrinsic sequential activity are thought to underlie motor sequencing<sup>2</sup>, navigation<sup>3,4</sup>, movement planning<sup>5</sup>, sensitivity to the timing of sensory stimuli<sup>6</sup> and cognitive tasks<sup>7</sup>. With few exceptions<sup>8</sup>, however, the biophysical mechanisms by which neural circuits produce sequences are poorly understood.

Songbirds have emerged as an excellent model system for investigating the neural mechanisms of sequence generation. The adult zebra finch song motif consists of a stereotyped pattern of song syllables<sup>9</sup>. One premotor forebrain area in particular, nucleus HVC (used as a proper name), is known to have a central role in controlling the temporal structure of birdsong<sup>10–12</sup>. During singing, neurons in HVC projecting to downstream premotor nucleus RA (robust nucleus of the arcopallium) produce only a single highly stereotyped burst of spikes during each repetition of the song motif<sup>13</sup>. Different RA-projecting HVC neurons (HVC<sub>(RA)</sub>) burst at different time points in the song, indicating that HVC neurons may burst sequentially through the song motif, in turn activating a complex and highly stereotyped pattern of bursts in the downstream nucleus RA<sup>14,15</sup>.

Here we set out to distinguish experimentally among several distinct classes of possible sequence-generating circuits within HVC. First, it has been proposed that sequential states of neural activity may be generated by synaptically connected chains of neurons<sup>6,16,17</sup>. In this view, activity could propagate through the HVC network—like a chain of falling dominoes—forming the basic clock that underlies song timing (Fig. 1a)<sup>10,18–20</sup>. A second, fundamentally different, class of models can allow for sequence generation in the absence of overt feed-forward connections between HVC<sub>(RA)</sub> neurons. In these models, oscillatory or other subthreshold dynamics can modulate the excitability of neurons and thus control the timing of their activity<sup>21,22</sup>, like those proposed to control the sequential activation of spikes during hippocampal theta sequences<sup>23</sup> and within replay events<sup>3,24</sup>. Subthreshold dynamics and rhythmicity on the timescale of song syllables (~100 ms) exist within HVC *in vitro*<sup>25</sup> and thus could have a central

role in controlling the timing of HVC bursts on that timescale in the singing bird.



**Figure 1 | Two broad classes of models for a sequence-generating circuit.** **a**, Neurons might form a feed-forward synaptically connected chain within the HVC such that activity propagates from one group of neurons to the next. **b**, Alternatively, sequential activity might occur in the absence of directed connections between neurons, from temporal and spatial gradients of excitability. For example, the network could receive a global and gradual ramping-down of an inhibitory input over time (red synapses), producing a sequential activation. The order of activation would be determined by neuronal excitability. In the example model shown here, neurons receive different levels of constant excitatory input (green synapses). The neuron with the largest excitatory input (neuron 1) would be most depolarized and would be the first to reach spiking threshold. The neuron with the smallest constant excitatory input (neuron 8) would be the last to reach threshold. In the model depicted here, the timescale of the sequence produced corresponds to one song syllable (shown above).

<sup>1</sup>McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

<sup>2</sup>Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. †Present address: Departments of Otolaryngology and Physiology and Neuroscience, NYU School of Medicine, 522 First Avenue, New York, New York 10016, USA.



## Intracellular recording during singing

To examine the role of subthreshold dynamics in the control of timing of HVC bursts during singing, we adopted an approach recently introduced for intracellular recordings in the freely moving rat<sup>26</sup>. We developed a miniature (1.6 g) microdrive that allows sharp microelectrode recordings to be performed in singing male zebra finches (Fig. 2a). Birds could move freely in a recording chamber, unrestrained except for a thin, flexible tether. In total, 28 neurons in 12 birds were recorded during singing of all three HVC neuron types, defined broadly by their axonal projections<sup>27,28</sup> (Fig. 2b).

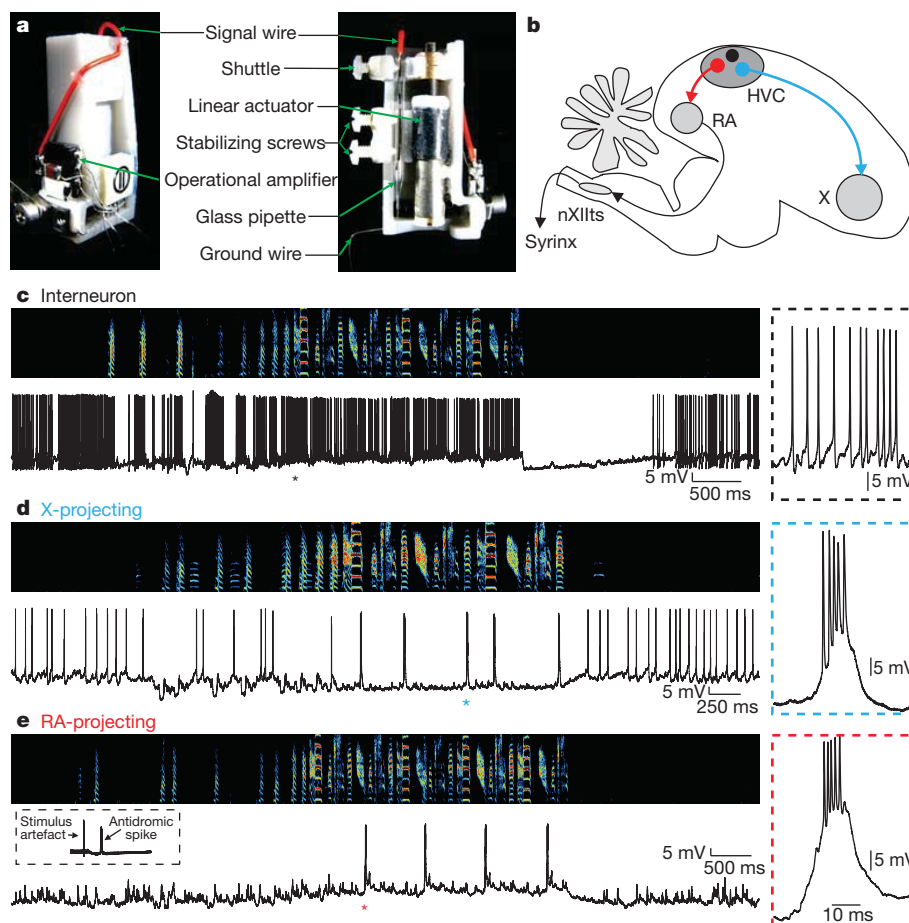
The singing-related spiking patterns of intracellularly recorded neurons closely resembled the previously described patterns in extracellular recordings<sup>13,29</sup>. Putative interneurons ( $n = 3$ ) were identified by a high spontaneous firing rate, and a continuous high firing rate throughout song (Fig. 2c,  $117 \pm 24.6$  Hz singing,  $66.3 \pm 21.6$  Hz baseline, error bars indicate  $\pm$  s.e.m. unless otherwise noted). Putative HVC neurons projecting to the basal ganglia homologue area X ( $n = 12$ ) exhibited a low spontaneous spiking rate ( $<10$  Hz) when the bird was not singing, and one or more high-frequency bursts during singing (Fig. 2d). These neurons showed a gradual hyperpolarization during the introductory notes (before the first motif in a bout of singing), and were hyperpolarized during song motifs (Fig. 2d,  $n = 12$  of 12 cells; singing,  $-70.8 \pm 3.4$  mV; baseline,  $-67.7 \pm 3.1$  mV), similar to what has been observed during auditory song playback<sup>30,31</sup>. We did not consider these neurons further in the context of sequence generation because it has been shown that selective ablation of X-projecting HVC neurons in adult zebra finches does not impair song production<sup>32</sup>.

HVC neurons that project to RA were identified by antidromic stimulation from RA (Fig. 2e, inset; see also Supplementary Fig. 1)<sup>13</sup>. HVC<sub>(RA)</sub> neurons showed a gradual depolarization before the onset of singing (Fig. 2e) and were persistently depolarized during singing ( $n = 13$  of 13 cells; singing,  $-67.3 \pm 3.5$  mV; baseline,  $-75.7 \pm 3.5$  mV). About half of HVC<sub>(RA)</sub> neurons ( $n = 7$  of 13) generated a single burst during each song motif (Fig. 3a–c,  $3.8 \pm 0.6$  spikes per burst). The remaining HVC<sub>(RA)</sub> neurons ( $n = 6$  of 13 cells) did not spike during song motifs (for example, Fig. 3d)<sup>13</sup>.

## Chain model versus ramp-to-threshold model

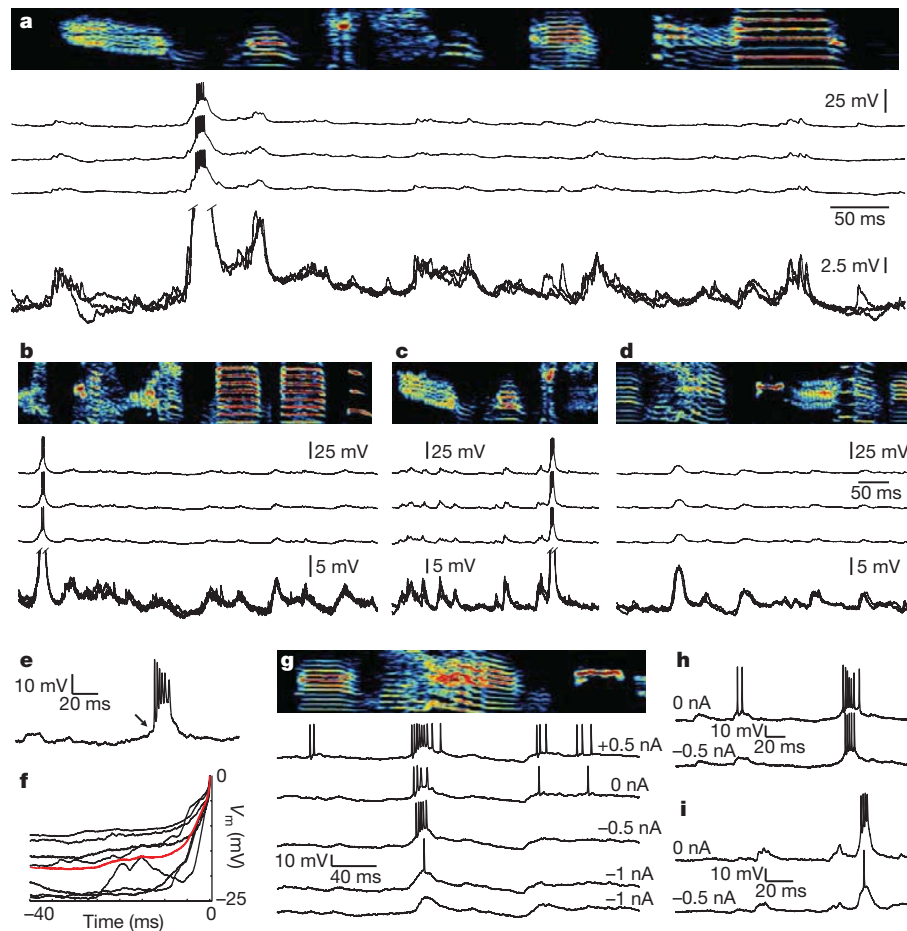
Recurrent synaptic connections within a network of sequentially active neurons would be expected to produce patterned synaptic inputs; thus previous reports of patterned synaptic inputs have been used as evidence of synaptically connected chains both *in vitro* and *in vivo*<sup>33</sup>. Consistent with this view, we observed a highly stereotyped pattern of fast subthreshold fluctuations widely distributed throughout the song (Figs 2e and 3a–d, and Supplementary Fig. 2). For individual neurons, the song-aligned subthreshold fluctuations were highly correlated across song motifs (cross-correlation  $0.80 \pm 0.04$ ,  $P < 10^{-9}$ ,  $n = 13$  neurons).

We now ask whether, as predicted by the ramp-to-threshold model, there was any slow ramping of membrane potential before the onset of bursts (Fig. 1b). We first consider the time window from the beginning of the song motif to the burst onset for each neuron. Across all 7 HVC<sub>(RA)</sub> neurons that burst during singing, the membrane potential did not change significantly in the period from the beginning of the



**Figure 2 | A microdrive for sharp intracellular recording in the singing bird.** **a**, The intracellular microdrive incorporates a motor that rotates a threaded rod and advances a shuttle that holds the electrode. **b**, A schematic of the zebra finch brain, highlighting three cell types in nucleus HVC defined by their projections: local circuit interneurons (in black), neurons that project to RA (in red), and

neurons that project to basal-ganglia-homologue area X (in blue). **c–e**, Examples of intracellular records from a putative local circuit interneuron (**c**), a putative X-projecting neuron (**d**) and an antidromically identified RA-projecting neuron (**e**). Asterisk indicates the region magnified in the panels to the right.



**Figure 3 | Intracellular membrane potential of identified  $HVC_{(RA)}$  neurons during singing.** **a–d**, Examples of the membrane potential of four  $HVC_{(RA)}$  neurons recorded during singing. For each cell, activity from three motif renditions is shown aligned to the song (top). Also shown is an overlay of the membrane potential traces (expanded vertical scale, bottom of each panel). **e**, Expanded view of a burst from another neuron during singing showing the flat membrane potential before burst onset (arrow). **f**, Average membrane

song motif to the moment 10 ms before the first spike in the burst ( $-0.47 \pm 0.69$  mV,  $P = 0.53$ ,  $t$ -test, average window duration,  $387 \pm 92$  ms). We next considered a ramp of excitation on the shorter timescale of a song syllable ( $\sim 100$  ms). Across all bursting neurons ( $n = 7$ ), the membrane potential did not change during a window from 100 ms to 10 ms before the first spike in the burst ( $0.31 \pm 1.04$  mV,  $P = 0.77$ ,  $t$ -test). Both of these results are inconsistent with a slow ramp of excitation before burst onset, on the timescale of either a song motif or a song syllable. In contrast, bursts of  $HVC_{(RA)}$  neurons were preceded, within the 5 ms before the first spike in the burst, by a large depolarization of  $10.5 \pm 1.9$  mV from baseline (Fig. 3e, f, the first spike of the burst initiated at a membrane potential of  $-52.6 \pm 1.7$  mV). This result is consistent with a model in which  $HVC_{(RA)}$  neurons are activated by a large synchronous synaptic input from a group of previously active neurons.

The two models described in Fig. 1 give very different predictions for the effect of intracellular current injection on the timing of neural activity. In a model in which the timing of  $HVC_{(RA)}$  bursts is controlled by slow membrane potential dynamics (Fig. 1b), an injected depolarizing current would cause the neuron to burst earlier during the slow depolarizing ramp, assuming that the burst-generating mechanism is sufficiently well coupled to the site of current injection (see Supplementary Discussion). In contrast, in the chain model, burst timing is controlled by a synaptic input from a preceding group of neurons (Fig. 1a). Thus, current injection would have a minimal

effect on burst timing, perhaps causing the first spike in the burst to appear a few milliseconds earlier during the onset of the synaptic depolarization. We assessed the effect of intracellular current injection on the timing of bursts in  $HVC_{(RA)}$  neurons during singing in three neurons. Two neurons were recorded with zero holding current and with 0.5 nA of hyperpolarizing current. One additional neuron was held long enough to record at four levels of holding current (0.5 nA, 0 nA,  $-0.5$  nA and  $-1$  nA). On average, the resulting membrane potential change was  $20.3$  mV nA $^{-1}$  of injected current. In all cases, hyperpolarizing current was seen to reduce the number of spikes in the burst (Fig. 3g–i, average 5 spikes per burst at 0 nA compared to 3.3 spikes per burst at  $-0.5$  nA), and could suppress spiking completely at the most hyperpolarizing currents ( $-1.0$  nA). Depolarizing current injection increased the number of spikes per burst (Fig. 3g).

Remarkably, the timing of the burst was only weakly affected by injected currents. At a hyperpolarizing holding current of 0.5 nA, the burst onset was delayed by an average of only 2.6 ms ( $n = 3$ ). However, the last spike of the burst was advanced by a similar amount such that the centre of the burst (midpoint between first and last spikes) was very weakly affected by injected current ( $1.2$  ms nA $^{-1}$ , Fig. 3g–i). In addition, under conditions at which the spiking was suppressed or nearly suppressed by hyperpolarizing current, a large underlying depolarization at the temporal position of the burst was clearly visible (Fig. 3g, i). These results are consistent with a mechanism in which a given



HVC<sub>(RA)</sub> neuron is driven by fast synaptic input from a preceding group of neurons.

### Cellular mechanisms of burst generation

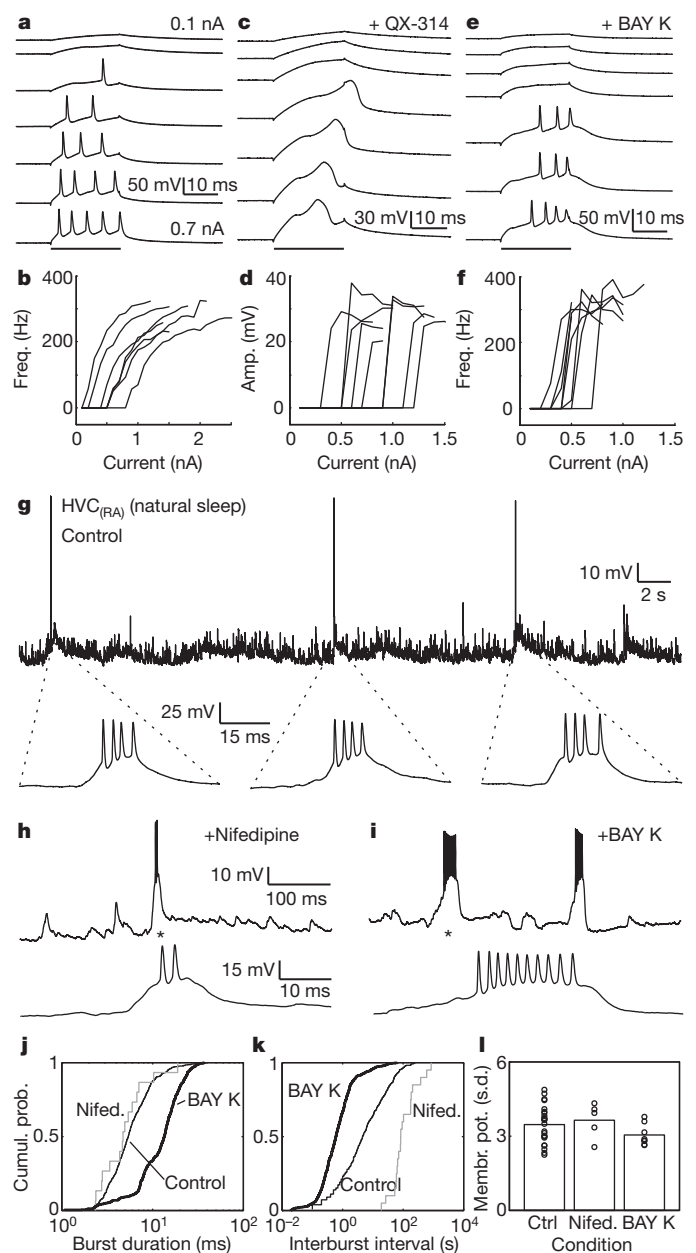
The broad powerful depolarizations that underlie the bursts of spikes in HVC<sub>(RA)</sub> neurons during singing (Fig. 3) are reminiscent of dendritic calcium spikes observed in many neurons<sup>34,35</sup>. Although it is difficult to establish definitively that the singing-related bursts of HVC<sub>(RA)</sub> neurons are mediated by calcium spikes, we have carried out *in vitro* and *in vivo* whole-cell recordings and pharmacological manipulations that support this view.

Although HVC<sub>(RA)</sub> neurons have not been observed to generate a burst response to somatic intracellular current injection (Fig. 4a, b)<sup>27,28,36</sup>, dendritic calcium spikes in some neurons may not be observed during somatic current injection<sup>37</sup>, but can be unmasked by the intracellular blockade of sodium and potassium channels<sup>38</sup>. We carried out whole-cell recordings in brain slices of antidromically identified HVC<sub>(RA)</sub> neurons with QX-314 in the recording pipette. Indeed, current injection resulted in a large depolarizing event in all neurons tested ( $n = 23$  cells, Fig. 4c, average amplitude  $26.4 \pm 5.6$  mV, width at half height  $4.5 \pm 1.0$  ms). The depolarizing events had a clear all-or-none response with an initiation threshold at the soma of  $-36.2 \pm 4.4$  mV (Fig. 4d,  $n = 14$  cells, compared to a threshold of  $-40.3 \pm 4.3$  mV for sodium spikes). In contrast, neurons in nucleus RA did not exhibit all-or-none spikes in the presence of QX-314 (ref. 39; Supplementary Fig. 3). The depolarizing events in HVC<sub>(RA)</sub> neurons were completely blocked by the broad spectrum calcium channel antagonist cadmium ( $100 \mu\text{M}$ ,  $n = 4$  cells), but were unaffected by nickel ( $100 \mu\text{M}$ ,  $n = 5$  cells), an antagonist of low-threshold voltage-gated calcium channels, indicating that the depolarizing events might be mediated by a high-threshold calcium channel.

We found that the L-type calcium channel agonist BAY K 8644 could enhance the calcium current sufficiently to evoke a burst response in HVC<sub>(RA)</sub> neurons even in the absence of QX-314 ( $n = 8$  cells, Fig. 4e, f, average of  $3.4 \pm 0.2$  spikes, within-burst spike rate  $302 \pm 14$  Hz). These burst responses appeared to have an all-or-none characteristic with a well-defined threshold for injected current ( $0.50 \pm 0.05$  nA), and a spike rate within bursts that did not increase at higher currents ( $P = 0.60$ ). These *in vitro* experiments indicate that HVC<sub>(RA)</sub> neurons are capable, under some conditions, of generating calcium-based regenerative spikes, possibly mediated by an L-type Ca conductance.

We wanted to examine more directly the role of these calcium conductances under conditions in which HVC<sub>(RA)</sub> neurons naturally generate burst sequences, rather than in brain slice. In a form of 'replay' of song-like patterns<sup>40</sup>, HVC<sub>(RA)</sub> and RA neurons generate sparse sequential bursts during sleep similar to those produced during singing<sup>13,41</sup>. We have adapted a head-fixed sleeping bird preparation<sup>13</sup> and used whole-cell recordings and pharmacological manipulation of HVC<sub>(RA)</sub> neurons to study the mechanisms underlying these bursts in naturally sleeping zebra finches (Fig. 4g). Across the population of HVC<sub>(RA)</sub> neurons in our data set ( $n = 36$  cells), nearly half the spikes recorded ( $49.3 \pm 3.5\%$ ) formed high-frequency bursts ( $>100$  Hz) during sleep ( $2.74 \pm 0.11$  sodium spikes, average within-burst rate of  $265 \pm 13$  Hz). Just as during singing, sleep bursts were seen to ride on a prominent underlying depolarizing event (Fig. 4g,  $25.2 \pm 0.9$  mV amplitude,  $18.4 \pm 1.5$  ms width at 2/3 height).

Injections of the L-type calcium channel agonist BAY K 8644 ( $100 \mu\text{M}$ , 5–20 nl bolus) in the vicinity ( $<100 \mu\text{m}$ ) of the whole-cell recording pipette increased the burst size (Fig. 4i, increased number of spikes and total burst duration,  $P < 10^{-5}$  for both measures, Kolmogorov-Smirnov test). In addition, these injections significantly increased the incidence of bursting (Fig. 4k, mean interburst interval  $2.0 \pm 5.7$  s with BAY K 8644, compared to  $18.4 \pm 34.5$  s control,  $P < 10^{-4}$ , Kolmogorov-Smirnov test,  $n = 6$  cells from 5 birds, mean  $\pm$  s.d.). In contrast, injections of the L-type calcium channel antagonist nifedipine ( $100 \mu\text{M}$ ) significantly decreased burst incidence (Fig. 4k,



**Figure 4 | Evidence that calcium channels contribute to burst events in HVC<sub>(RA)</sub> neurons.** **a**, Response of an HVC<sub>(RA)</sub> neuron in brain slice to somatically injected current steps (black bar) of different size. **b**, Relationship between injected current and evoked firing rate in a population of 7 HVC<sub>(RA)</sub> neurons. Note that somatic current injection does not elicit an all-or-none burst. **c**, In the presence of intracellular sodium and potassium channel blocker QX-314 (5 mM), calcium spikes appear as an all-or-none depolarizing event. **d**, The amplitude of the depolarizing event (threshold to maximum point) as a function of injected current reveals an all-or-none response ( $n = 8$  of 8 cells). **e**, **f**, HVC<sub>(RA)</sub> neurons treated with the L-type calcium channel agonist BAY K 8644 (5–10  $\mu\text{M}$ ) generate all-or-none spike bursts in response to somatic current injection. **g**, Segment of a whole-cell recording in a head-fixed bird during natural sleep showing three spontaneous bursts. **h**, **i**, Spontaneous bursting activity recorded during sleep after localized injection of L-type calcium channel antagonist nifedipine (**h**) or agonist BAY K 8644 (**i**). Asterisk indicates expanded view below. **j**, **k**, Cumulative distribution of burst durations and inter-burst intervals for control, nifedipine and BAY K 8644 conditions. **l**, Standard deviation of membrane potential fluctuations is not affected by nifedipine or BAY K 8644, indicating that synaptic transmission is not affected by these drugs.

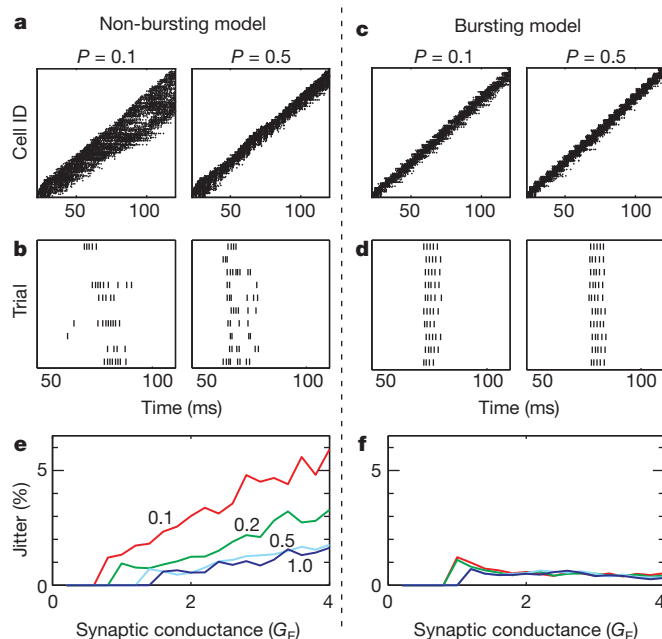
mean interburst interval  $171.7 \pm 209.6$  s, greater than control,  $P < 0.0001$ , Kolmogorov-Smirnov test,  $n = 6$  cells from 4 birds, mean  $\pm$  s.d.). The effect of L-type calcium channel modulators could

not be explained by changes in the size of synaptic inputs: the magnitude of fluctuations in membrane potential was not altered by BAY K 8644 or nifedipine ( $P > 0.05$ ,  $t$ -test, Fig. 4l). Taken together, these experiments demonstrate that L-type calcium channels have a role in generating or initiating bursting activity in  $HVC_{(RA)}$  neurons. Such highly nonlinear all-or-none calcium spikes produce a highly stereotyped response to a wide range of synaptic inputs<sup>42</sup>, and could have implications for the propagation of activity in a synaptically connected chain of neurons.

### Burst propagation in a chain network

The stable propagation of bursts in an excitatory chain network is non-trivial; it requires precisely tuned synaptic strengths to avoid runaway excitation or decay<sup>19</sup>. It has previously been shown that an intrinsic neuronal burst mechanism can allow the stable propagation of activity in a chain network<sup>19</sup>, but what about temporal precision and stereotypy? Here we use a simple biophysical model to examine the role that intrinsic bursting might have in achieving precise stereotyped temporal structure in the presence of noise. We also examine how such a mechanism might make the functioning of these networks robust over a wide range of network and synaptic properties.

We studied a network of 70 groups of 30 excitatory  $HVC_{(RA)}$  neurons each, organized in a sequentially connected chain. Recurrent inhibition in  $HVC_{(RA)}$ <sup>25,43</sup> was implemented by a population of 300 interneurons with sparse random connections to the excitatory chain (Supplementary Fig. 4a). We began with a non-bursting model of  $HVC_{(RA)}$  neurons, described by a single spiking somatic compartment (Fig. 5, Supplementary Fig. 4b and Supplementary Methods).



**Figure 5 | A simple biophysical model to examine the implications of neuronal bursting on the robustness of HVC network propagation.** **a–d**, Two models of a synaptically connected chain network were compared: one with non-bursting neurons (**a**, **b**), the other with bursting neurons (**c**, **d**). **a**, Non-bursting model: spike raster plot for all neurons in the network showing activity as a function of time for two different levels of network connection probability ( $P = 0.1$  and  $0.5$ ). **b**, Spike raster of a single neuron during different runs of the network. Note the non-stationarity of propagation and large variability across runs. **c**, Bursting model: spike raster plot for all neurons in the network. **d**, Spike raster of a single neuron during different runs of the network. Note the highly uniform propagation and stereotyped response across runs. **e**, **f**, Run-time jitter, plotted as a function of network connectivity and synaptic conductance, is consistently lower in the bursting model than in the non-bursting model. (See Supplementary Figures and Table for further quantification, and Supplementary Methods for model details.)

We found that this network did not exhibit the unstable (explosive or decaying) behaviour characteristic of purely excitatory networks<sup>19</sup>, but exhibited stable propagation of burst activity over a wide range of connection probabilities ( $P = 0.1$ – $1.0$ ) and excitatory synaptic strengths between  $HVC_{(RA)}$  neurons in successive groups ( $G_{EE,max}$  from  $0.2$  to  $4.0 \text{ mS cm}^{-2}$ ). Nevertheless, the activity tended to be non-stationary, particularly at lower connection probabilities ( $P = 0.1$ , Fig. 5a), exhibiting both dispersion (broadening) and variations in propagation velocity at different points in the network (Supplementary Figs 5 and 6a, b). Furthermore, the network was sensitive to the presence of noise, producing activity that was not stereotyped across multiple trials of the simulation, including large jitter in the speed of propagation through the network (Fig. 5b, e;  $1.95 \pm 1.38\%$  mean run-time jitter  $\pm$  s.d.) and large variations in the burst response on different trials (quantified as spikes per burst and burst unreliability, Supplementary Fig. 6c–e). Finally, many characteristics of the propagation (number of spikes per burst, burst duration and burst jitter) were strongly dependent on the network connection probabilities and connection strengths (Fig. 5 and Supplementary Fig. 6). Thus, although the stable propagation of bursts is possible in a chain network of non-bursting neurons, the network does not produce the stereotyped sequences characteristic of real  $HVC_{(RA)}$  neurons.

The situation was markedly different in a model with neurons that have an intrinsic burst mechanism. Bursting  $HVC_{(RA)}$  neurons were modelled with a spiking somatic compartment plus a dendritic compartment containing conductances for generating calcium spikes (see Supplementary Figs 4c, d and 7–9). Propagation down the chain was stationary, with no broadening or variations in velocity (Fig. 5c and Supplementary Fig. 5). The propagation was also extremely stereotyped, exhibiting small trial-to-trial variations in propagation speed (Fig. 5d,  $0.52 \pm 0.17\%$  mean run-time jitter  $\pm$  s.d.). Burst response was much more reliable in the bursting model (see spikes per burst and burst unreliability, Supplementary Fig. 6), similar to what has been observed in singing-related firing patterns of  $HVC_{(RA)}$  neurons<sup>13,29</sup>. Finally, in the bursting model, every characteristic of burst propagation that we examined was much more robust to variations in network connection probability and synaptic strength than was the single compartment model (Fig. 5e, f and Supplementary Fig. 6). Similar results were obtained with a simple integrate-and-burst model (Supplementary Fig. 10). Taken together, these results indicate that an intrinsic neuronal burst mechanism, regardless of its biophysical implementation, could serve a fundamental role in allowing synaptically connected chain networks to propagate in a highly stereotyped manner with low temporal jitter, even in the presence of noise, and over a wide range of network connectivities. Such robustness could also make sequence-generating networks easier to assemble during development<sup>44,45</sup>.

We have carried out intracellular recording and manipulation of activity in the freely behaving animal in a neural circuit important for the temporal control of behaviour. We observed no ramping or rhythmicity that could contribute to the temporal patterning of  $HVC_{(RA)}$  bursts. In contrast, our recordings reveal a single large postsynaptic potential that immediately precedes the onset of a song-locked burst of spikes. Together, our findings are consistent with the idea that the control of song temporal structure is produced by the propagation of calcium-mediated bursts through a synaptically connected chain of neurons. Temporally precise learned behaviours in other vertebrates could use similar mechanisms to organize neuronal activity into sequentially active states.

### METHODS SUMMARY

**Subjects.** We used adult ( $> 120$  days post hatch) male zebra finches (*Taeniopygia guttata*). All animal procedures were reviewed and approved by the MIT committee on animal care.

**Intracellular recording during singing.** Intracellular recordings were achieved in the zebra finch using a custom microdrive constructed out of 3D printed plastic (AP Proto) outfitted with a lightweight linear actuator (Smooovy Series 0515,



Faulhaber). A preamplifier was mounted at the base of the device which routed signals to a commercially available intracellular amplifier (IR-183, Cygnus Technology). Sharp microelectrodes were pulled to a final impedance of 80–110 M $\Omega$  and were filled with 3 M potassium acetate. Once a stable intracellular recording was obtained, a female bird was presented to elicit directed singing.

**Intracellular recording during sleep.** During an initial surgical step, a stainless steel headplate was affixed to the skull. A small (~200  $\mu$ m) craniotomy was made over HVC. Whole-cell recordings were made with glass electrodes (5–8 M $\Omega$ ) using techniques described elsewhere<sup>46</sup>. Signals were measured using an Axoclamp 2B (Molecular Devices). In some experiments, an injection pipette (20–30  $\mu$ m opening) was positioned less than 100  $\mu$ m from the recording site for the injection (Nanoject II, Drummond Scientific) of a small volume (5–20 nl) of 100  $\mu$ M (+/–) BAY K 8644 (A.G. Scientific) or 100  $\mu$ M nifedipine (Sigma).

**Slice preparation.** 400- $\mu$ m slices were prepared on a vibrating microtome (Leica VT1000) and placed in ice-cold ACSF (sodium replaced with equimolar sucrose). Slices were then recorded in an interface-style chamber (VB5000, Leica) with standard ACSF (in mM): 126 NaCl, 3 KCl, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, 2 MgSO<sub>4</sub>·7H<sub>2</sub>O, 26 NaHCO<sub>3</sub>, 10 dextrose, 2 CaCl<sub>2</sub>·2H<sub>2</sub>O. QX-314 (5 mM, internal) was used in a subset of these experiments.

Received 16 May; accepted 2 September 2010.

Published online 24 October 2010.

- Lashley, K. In *Cerebral Mechanisms in Behavior* (ed. Jeffress, L.) (Wiley, 1951).
- Shima, K., Isoda, M., Mushiake, H. & Tanji, J. Categorization of behavioural sequences in the prefrontal cortex. *Nature* **445**, 315–318 (2007).
- Mehta, M. R., Lee, A. K. & Wilson, M. A. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741–746 (2002).
- Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
- Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsaki, G. Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327 (2008).
- Mauk, M. D. & Buonomano, D. V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
- Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B. & Massey, J. T. Mental rotation of the neuronal population vector. *Science* **243**, 234–236 (1989).
- Stent, G. S. et al. Neuronal generation of the leech swimming movement. *Science* **200**, 1348–1357 (1978).
- Konishi, M. Birdsong: from behavior to neuron. *Annu. Rev. Neurosci.* **8**, 125–170 (1985).
- Long, M. A. & Fee, M. S. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194 (2008).
- Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871–1875 (1996).
- Vu, E. T., Mazurek, M. E. & Kuo, Y. C. Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* **14**, 6924–6934 (1994).
- Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
- Leonardo, A. & Fee, M. S. Ensemble coding of vocal control in birdsong. *J. Neurosci.* **25**, 652–661 (2005).
- Fee, M. S., Kozhevnikov, A. A. & Hahnloser, R. H. Neural mechanisms of vocal sequence generation in the songbird. *Ann. NY Acad. Sci.* **1016**, 153–170 (2004).
- Amari, S. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comp.* **c-21**, 1197–1206 (1972).
- Abeles, M. *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge Univ. Press, 1991).
- Li, M. & Greenside, H. Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. *Phys. Rev. E* **74**, 011918 (2006).
- Jin, D. Z., Ramazanoglu, F. M. & Seung, H. S. Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. *J. Comput. Neurosci.* **23**, 283–299 (2007).
- Glaze, C. M. & Troyer, T. W. Behavioral measurements of a temporally precise motor code for birdsong. *J. Neurosci.* **27**, 7631–7639 (2007).
- Mann, E. O. & Paulsen, O. Role of GABAergic inhibition in hippocampal network oscillations. *Trends Neurosci.* **30**, 343–349 (2007).
- O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
- Foster, D. J. & Wilson, M. A. Hippocampal theta sequences. *Hippocampus* **17**, 1093–1099 (2007).
- Buzsaki, G. Two-stage model of memory trace formation: a role for “noisy” brain states. *Neuroscience* **31**, 551–570 (1989).
- Solis, M. M. & Perkel, D. J. Rhythmic activity in a forebrain vocal control nucleus *in vitro*. *J. Neurosci.* **25**, 2811–2822 (2005).
- Lee, A. K., Manns, I. D., Sakmann, B. & Brecht, M. Whole-cell recordings in freely moving rats. *Neuron* **51**, 399–407 (2006).
- Dutar, P., Vu, H. M. & Perkel, D. J. Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVC of the adult zebra finch. *J. Neurophysiol.* **80**, 1828–1838 (1998).
- Mooney, R. Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J. Neurosci.* **20**, 5420–5436 (2000).
- Kozhevnikov, A. A. & Fee, M. S. Singing-related activity of identified HVC neurons in the zebra finch. *J. Neurophysiol.* **97**, 4271–4283 (2007).
- Lewicki, M. S. Intracellular characterization of song-specific neurons in the zebra finch auditory forebrain. *J. Neurosci.* **16**, 5855–5863 (1996).
- Rosen, M. J. & Mooney, R. Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus HVC. *Neuron* **39**, 177–194 (2003).
- Scharff, C., Kirn, J. R., Grossman, M., Macklis, J. D. & Nottebohm, F. Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. *Neuron* **25**, 481–492 (2000).
- Ikegaya, Y. et al. Synfire chains and cortical songs: temporal modules of cortical activity. *Science* **304**, 559–564 (2004).
- London, M. & Hausser, M. Dendritic computation. *Annu. Rev. Neurosci.* **28**, 503–532 (2005).
- Murayama, M. et al. Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. *Nature* **457**, 1137–1141 (2009).
- Kubota, M. & Taniguchi, I. Electrophysiological characteristics of classes of neuron in the HVC of the zebra finch. *J. Neurophysiol.* **80**, 914–923 (1998).
- Larkum, M. E. & Zhu, J. J. Signaling of layer 1 and whisker-evoked Ca<sup>2+</sup> and Na<sup>+</sup> action potentials in distal and terminal dendrites of rat neocortical pyramidal neurons *in vitro* and *in vivo*. *J. Neurosci.* **22**, 6991–7005 (2002).
- Connors, B. W. & Prince, D. A. Effects of local anesthetic QX-314 on the membrane properties of hippocampal pyramidal neurons. *J. Pharmacol. Exp. Ther.* **220**, 476–481 (1982).
- Mooney, R. Synaptic basis for developmental plasticity in a birdsong nucleus. *J. Neurosci.* **12**, 2464–2477 (1992).
- Dave, A. S. & Margoliash, D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290**, 812–816 (2000).
- Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. Sleep-related neural activity in a premotor and a basal-ganglia pathway of the songbird. *J. Neurophysiol.* **96**, 794–812 (2006).
- Wei, D. S. et al. Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science* **293**, 2272–2275 (2001).
- Mooney, R. & Prather, J. F. The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J. Neurosci.* **25**, 1952–1964 (2005).
- Jun, J. K. & Jin, D. Z. Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. *PLoS ONE* **2**, e723 (2007).
- Fiete, I. R., Senn, W., Wang, C. Z. & Hahnloser, R. H. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65**, 563–576 (2010).
- Margrie, T. W., Brecht, M. & Sakmann, B. *In vivo*, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Arch.* **444**, 491–498 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Wilson, S. Seung, A. Andalman, J. Goldberg and A. Gray for comments on earlier versions of this manuscript. We would also like to thank A. Andalman, D. Aronov and T. Ramee for help with acquisition and analysis software. This work is supported by funding from the National Institutes of Health to M.S.F. (MH067105) and M.A.L. (DC009280), and from the Alfred P. Sloan Research Fellowship and the National Science Foundation to D.Z.J. (IOS-0827731).

**Author Contributions** M.S.F. and M.A.L. conceived and designed the experiments and analysed the experimental data. M.A.L. acquired the experimental data. M.S.F., M.A.L. and D.Z.J. designed, and D.Z.J. carried out, the modelling experiments. All authors contributed to writing the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.S.F. (fee@mit.edu).

# Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor

Laura B. Sheard<sup>1\*</sup>, Xu Tan<sup>1\*†</sup>, Haibin Mao<sup>1\*</sup>, John Withers<sup>2,3</sup>, Gili Ben-Nissan<sup>4</sup>, Thomas R. Hinds<sup>1</sup>, Yuichi Kobayashi<sup>5</sup>, Fong-Fu Hsu<sup>6</sup>, Michal Sharon<sup>4</sup>, John Browse<sup>7</sup>, Sheng Yang He<sup>2,3</sup>, Josep Rizo<sup>8</sup>, Gregg A. Howe<sup>2,9</sup> & Ning Zheng<sup>1,10</sup>

Jasmonates are a family of plant hormones that regulate plant growth, development and responses to stress. The F-box protein CORONATINE INSENSITIVE 1 (COI1) mediates jasmonate signalling by promoting hormone-dependent ubiquitylation and degradation of transcriptional repressor JAZ proteins. Despite its importance, the mechanism of jasmonate perception remains unclear. Here we present structural and pharmacological data to show that the true *Arabidopsis* jasmonate receptor is a complex of both COI1 and JAZ. COI1 contains an open pocket that recognizes the bioactive hormone (3R,7S)-jasmonoyl-L-isoleucine (JA-Ile) with high specificity. High-affinity hormone binding requires a bipartite JAZ degron sequence consisting of a conserved  $\alpha$ -helix for COI1 docking and a loop region to trap the hormone in its binding pocket. In addition, we identify a third critical component of the jasmonate co-receptor complex, inositol pentakisphosphate, which interacts with both COI1 and JAZ adjacent to the ligand. Our results unravel the mechanism of jasmonate perception and highlight the ability of F-box proteins to evolve as multi-component signalling hubs.

The phytohormone jasmonate and its metabolites regulate a wide spectrum of plant physiology, participating in normal development and growth processes, as well as defence responses to environmental and pathogenic stressors<sup>1</sup>. Jasmonate is activated upon specific conjugation to the amino acid L-isoleucine (Ile), which produces the highly bioactive hormonal signal (3R,7S)-jasmonoyl-L-isoleucine (JA-Ile) that is functionally and structurally mimicked by the *Pseudomonas syringae* phytotoxin coronatine<sup>2–4</sup>. The discovery of coronatine-insensitive mutants enabled the identification of COI1 as a key player in the jasmonate pathway, with further implications of regulated proteolysis in jasmonate perception and signal transduction<sup>5</sup>.

COI1 is an F-box protein that functions as the substrate-recruiting module of the Skp1-Cul1-F-box protein (SCF) ubiquitin E3 ligase complex. Recent studies have identified the JASMONATE ZIM DOMAIN (JAZ) family of transcriptional repressors as SCF<sup>COI1</sup> substrate targets, which associate with COI1 in a hormone-dependent manner<sup>6–8</sup>. In the absence of hormone signal, JAZ proteins actively repress the transcription factor MYC2, which binds to *cis*-acting elements of jasmonate-response genes. In response to cues that upregulate JA-Ile synthesis, the hormone stimulates the specific binding of JAZ proteins to COI1, leading to poly-ubiquitylation and subsequent degradation of JAZ by the 26S proteasome. JAZ degradation relieves repression of MYC2 and probably other transcription factors, permitting the expression of jasmonate-responsive genes<sup>6,9</sup>. The role of COI1-mediated JAZ degradation in jasmonate signalling is analogous to auxin signalling through the receptor F-box protein TIR1, which promotes hormone-dependent turnover of the AUX/IAA transcriptional repressors<sup>10,11</sup>. Supported by its

sequence homology and functional similarity to TIR1, COI1 has been assigned a critical role in the direct perception of the jasmonate signal<sup>12,13</sup>.

Despite the importance of jasmonate signalling in plant physiology, the molecular mechanism of jasmonate perception remains elusive. Here we present crystal structures of COI1 bound to JA-Ile or coronatine, as well as peptides of a bipartite JAZ1 degron. Our structural and pharmacological studies reveal that the true jasmonate receptor is a co-receptor complex, consisting of the F-box protein COI1, the JAZ degron and a newly discovered third component, inositol pentakisphosphate.

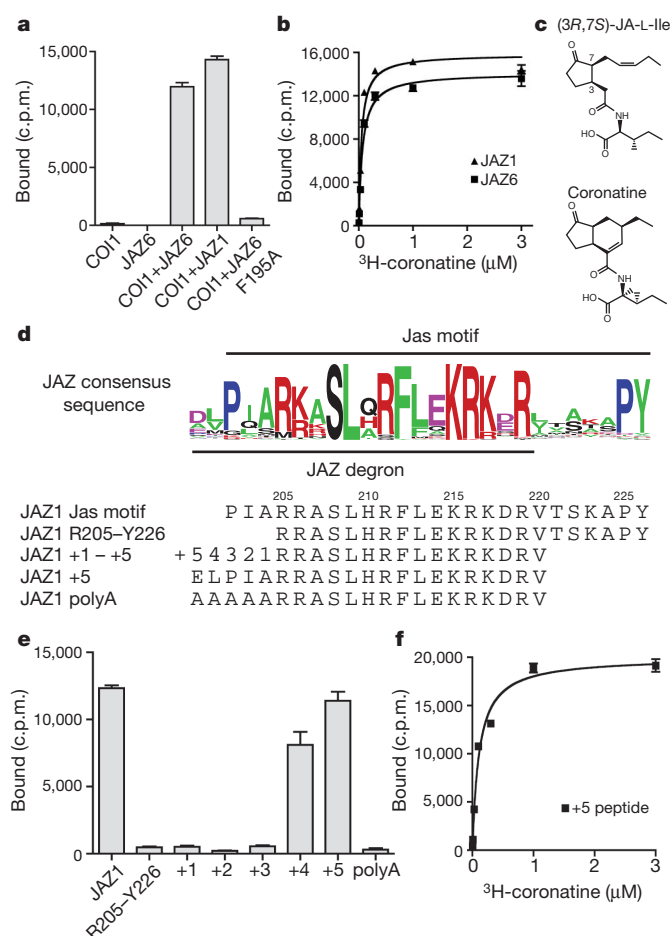
## COI1-JAZ complex as a jasmonate co-receptor

To characterize better the pharmacology of jasmonate perception, we used recombinant proteins and <sup>3</sup>H-coronatine to define quantitatively the functional components of the receptor system with an *in vitro* radioligand binding assay. From saturation binding experiments, we detected high-affinity specific binding of <sup>3</sup>H-coronatine to COI1 in the presence of two different full-length JAZ proteins (JAZ6 and JAZ1 at a dissociation constant ( $K_d$ ) of 68 nM and 48 nM, respectively; Fig. 1a, b). The highly active (3R,7S) isomer of JA-Ile (Fig. 1c) and the less active (3R,7R) isomer compete with <sup>3</sup>H-coronatine for binding to the COI1-JAZ6 complex with inhibition constant ( $K_i$ ) values of 1.8  $\mu$ M and 18  $\mu$ M, respectively (Supplementary Fig. 1a). In contrast, <sup>3</sup>H-coronatine displayed no affinity to the JAZ proteins and exhibited only marginal binding to the F-box protein alone. Hormone binding to COI1 alone elicited <2% binding signal relative to that of COI1-JAZ at a concentration that saturates the complex (300 nM) (Fig. 1a and Supplementary Fig. 1b). This result, together with the observation

<sup>1</sup>Department of Pharmacology, Box 357280, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824, USA. <sup>3</sup>Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, USA. <sup>4</sup>Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>5</sup>Department of Biological Engineering, Tokyo Institute of Technology, 4259-B52 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan. <sup>6</sup>Mass Spectrometry Resource, Division of Endocrinology, Diabetes, Metabolism, and Lipid research, Department of Internal Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>7</sup>Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164, USA. <sup>8</sup>Department of Biochemistry, University of Texas Southwestern Medical Center, 6000 Harry Hines Boulevard, Dallas, Texas 75390, USA. <sup>9</sup>Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA. <sup>10</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. <sup>†</sup>Present address: Division of Genetics, Brigham and Women's Hospital, Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

\*These authors contributed equally to this work.





**Figure 1 | COI1-ASK1 and JAZ proteins form a high-affinity jasmonate co-receptor.** **a**, Binding of tritium-labelled coronatine (300 nM) to recombinant COI1-ASK1 and JAZ proteins. c.p.m., counts per minute. **b**, Saturation binding of <sup>3</sup>H-coronatine to the complex of COI1-ASK1 in the presence of JAZ6 or JAZ1, with a  $K_d$  of  $68 \pm 15$  nM and  $48 \pm 13$  nM, respectively. **c**, Chemical structures of (3R,7S)-JA-Ile and coronatine. **d**, The consensus sequence of the Jas motif from 61 JAZ proteins from two monocotyledon and three dicotyledon plant species. Corresponding peptide sequences from JAZ1 in **e** are listed below. **e**, <sup>3</sup>H-coronatine binding at 300 nM to COI1 in the presence of a series of synthetic JAZ1 peptides with the N terminus of R205-Y226 systematically extended as described in **d**. **f**, Saturation binding of COI1-ASK1 and the JAZ1 +5 degron peptide, with a  $K_d$  of  $108 \pm 29$  nM. All results are the mean  $\pm$  s.e. of two to three experiments performed in duplicate.

that endogenous JA-Ile activates COI1-dependent gene expression in the nanomolar range<sup>1,14,15</sup>, indicates that the COI1-JAZ complex, rather than COI1 alone, functions as the genuine high-affinity jasmonate receptor in a co-receptor form.

We have previously mapped the COI1-binding region of the JAZ proteins to the carboxy-terminal Jas motif, which is characterized by the SLX<sub>2</sub>FX<sub>2</sub>KRX<sub>2</sub>RX<sub>5</sub>PY consensus sequence preceded by two consecutive basic residues<sup>16,17</sup> (Fig. 1d). A single Ala mutation of the central strictly conserved phenylalanine residue in the Jas motif is sufficient to abolish the formation of the high-affinity jasmonate co-receptor (Fig. 1a). Previous studies showed that the highly conserved PY sequence at the C terminus of the Jas motif has a role in JAZ localization and stability *in vivo*, but is not required for ligand-dependent COI1-JAZ interaction<sup>16,18,19</sup>. Consistent with these findings, truncation of the PY motif in JAZ1 has little effect on the *in vitro* ligand-binding activity (Supplementary Fig. 1c).

To map further the minimal region of the Jas motif required for high-affinity ligand binding with COI1, we replaced the recombinant protein with synthetic peptides of JAZ1 in the ligand binding assay (Fig. 1d, e). A 22-amino-acid JAZ1 peptide (Arg 205-Tyr 226) spanning the central

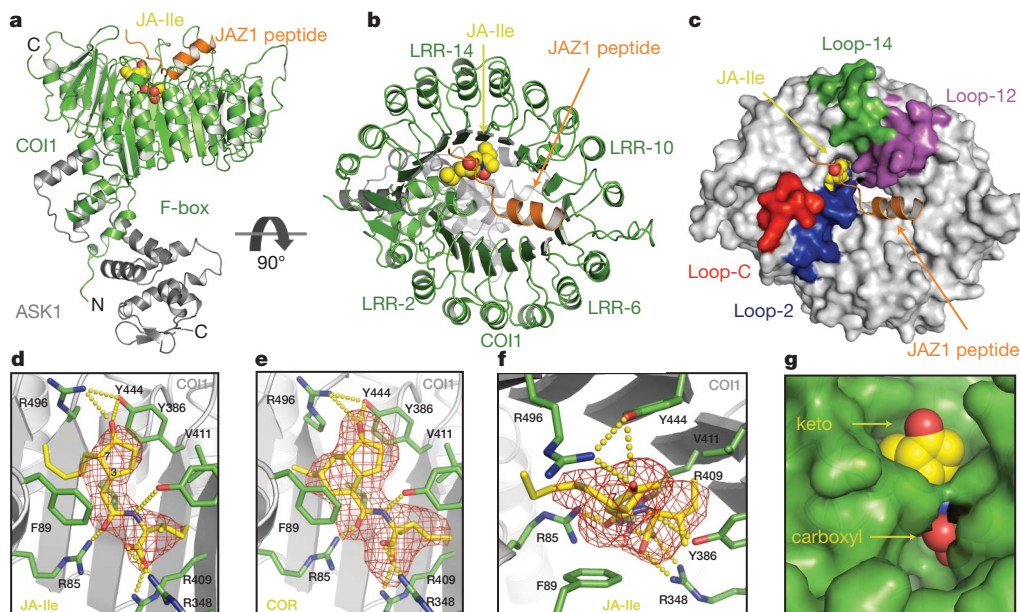
conserved Jas motif plus the two amino-terminal basic residues was not sufficient to form the high-affinity jasmonate co-receptor with COI1, indicating that amino acids N-terminal to Arg 205 also participate in the COI1-jasmonate interaction. Because several JAZ proteins show sequence homology in this region (Fig. 1d), we tested a series of JAZ1 peptides in which the N terminus was systematically extended by one amino acid. Notably, inclusion of four but not three amino acids N-terminal to Arg 205 allows ligand-dependent co-receptor formation, whereas addition of the fifth residue (Glu 200) to the JAZ1 peptide permits <sup>3</sup>H-coronatine binding with a  $K_d$  comparable to that of the full-length JAZ1 protein (Fig. 1e, f). Despite the sequence variation among different JAZ members in this region, only select amino acids are functional in this five-amino-acid extension, as a penta-alanine sequence fails to elicit the same effect (Fig. 1e). Together, these results indicate that the JAZ1 protein uses a minimal sequence (Glu 200-Val 220) within the Jas motif, which consists of a highly conserved central and C-terminal region and a more variable N-terminal region, to interact with COI1 and perceive the jasmonate signal. Consistent with our *in vitro* ligand-binding data, the minimal sequence in JAZ1 is sufficient for coronatine-induced COI1-JAZ1 interaction (Supplementary Fig. 1d). Therefore, we conclude that the interactions among COI1, coronatine and the JAZ1 peptide are highly cooperative and that the short Glu 200-Val 220 sequence functions as the JAZ1 degron.

### Jasmonate-binding pocket on COI1

To elucidate the structural mechanism by which the COI1-JAZ1 co-receptor senses jasmonate, we crystallized and determined the structures of the COI1-ASK1-JAZ1 degron peptide complex together with either (3R,7S)-JA-Ile or coronatine (Supplementary Table 1). The crystal structure of COI1 reveals a TIR1-like overall architecture<sup>20</sup>, with an N-terminal tri-helical F-box motif bound to ASK1 and a C-terminal horseshoe-shaped solenoid domain formed by 18 tandem leucine-rich repeats (LRRs; Fig. 2a, b). Similar to TIR1, the top surface of the COI1 LRR domain has three long intra-repeat loops (loop-2, loop-12 and loop-14) that are involved in hormone and polypeptide substrate binding. Unlike TIR1, however, a fourth long loop (loop-C) in the C-terminal capping sequence of the COI1 LRR domain folds over loop-2, partially covering it from above (Fig. 2b, c).

Despite their similar overall fold, COI1 has evolved a hormone-binding site that is distinct from TIR1. Configured in between loop-2 and the inner wall of the LRR solenoid, the ligand-binding pocket of COI1 is exclusively encircled by amino acid side chains (Fig. 2d-f). Many of the pocket-forming residues on COI1 are large in size and carry a polar head group (Supplementary Fig. 2). These properties allow them to mould a binding pocket into a specific shape while forming close interactions with each chemical moiety of the ligand. These close interactions are critical to proper hormone sensing of the complex—in yeast two-hybrid assays, mutation of any of these large side-chain amino acids on COI1 is sufficient to disrupt the interaction of COI1 with JAZ1 in the presence of coronatine (Supplementary Fig. 3).

In the binding pocket, both JA-Ile and coronatine sit in an 'upright' position with the keto group of their common cyclopentanone ring pointing up and forming a triangular hydrogen bond network with Arg 496 and Tyr 444 of COI1 at the pocket entrance (Fig. 2d-f). Without the JAZ degron peptide bound, the keto group of the ligand is accessible to solvent (Fig. 2g). The rest of the cyclopentanone ring of both JA-Ile and coronatine is sandwiched between the aromatic groups of Phe 89 and Tyr 444 of COI1, stabilized by hydrophobic packing. The cyclohexene ring of coronatine provides a rigid surface area for close packing with Phe 89, whereas the more flexible and extended pentenyl side chain of JA-Ile is more loosely accommodated by a hydrophobic pocket formed by Ala 86, Phe 89 and Leu 91 from loop-2 as well as Leu 469 and Trp 519 from the LRRs (Supplementary Fig. 4a). Differences at this interface probably explain the approximately tenfold higher affinity of coronatine over (3R,7S)-JA-Ile, as detected in our binding assays.



**Figure 2 | Crystal structure of the COI1-ASK1 complex with JA-Ile and the JAZ1 degron peptide.** **a, b**, COI1-ASK1 (green and grey ribbons, respectively) with the JAZ1 degron peptide (orange ribbon) and (3R,7S)-JA-Ile in yellow space-fill representation. **c**, Surface representation of COI1 (grey) with loop-2 (blue), loop-12 (purple) and loop-14 (green) forming the JA-Ile binding pocket. **d, e**, Side view of (3R,7S)-JA-Ile (JA-Ile) and coronatine (COR) binding. Hormones are shown as stick models, along with positive  $F_o - F_c$  electron density, calculated before they were built into the model (red mesh). Hydrogen

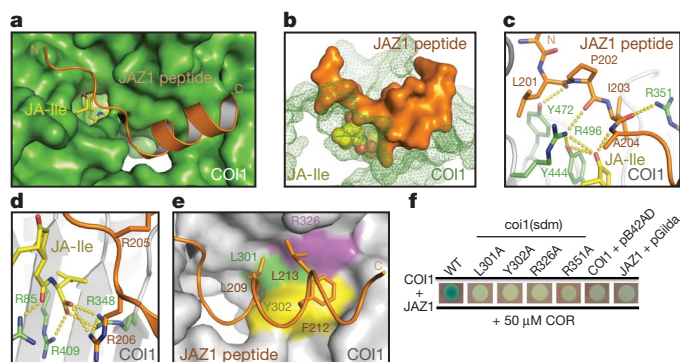
Deeper in the ligand-binding pocket, the common amide and carboxyl groups of JA-Ile and coronatine bind to the bottom of the binding site by forming a salt bridge and hydrogen bond network with three basic residues of COI1: Arg 85, Arg 348 and Arg 409 (Fig. 2d, e). Together, these arginine residues constitute the charged floor of the ligand pocket. Tyr 386 reinforces the interactions from above by making a hydrogen bond with the amine group of the ligand. In doing so, Tyr 386 approaches the cyclopentanone ring of the ligand, narrowing the pocket entrance, and creating a hydrophobic cave below. The rest of the basin is carved out by Val 411, Ala 384 and the aliphatic side chain of Arg 409 (Supplementary Fig. 4b). The ethyl-cyclopropane group of coronatine and the isoleucine side chain of JA-Ile can both comfortably fit in this space due to their similar size and hydrophobicity. The nature of the cave explains the preference of COI1 for jasmonate conjugates containing a moderately sized hydrophobic amino acid<sup>13</sup>. Although most of the ligand is buried inside the binding site, the keto group at the top and the carboxyl group at the bottom remain exposed, available for additional interactions with the JAZ1 portion of the co-receptor (Fig. 2g).

### Structural roles of the bipartite JAZ1 degron

The JAZ1 degron peptide adopts a bipartite structure with a loop region followed by an  $\alpha$ -helix to assemble with the COI1-jasmonate complex. The hallmark of the JAZ1 degron is the N-terminal five amino acids identified in the radioligand binding assay. In a largely extended conformation, this short sequence lies on top of the hormone-binding pocket and simultaneously interacts with both COI1 and the ligand, effectively trapping the ligand in the pocket (Fig. 3a, b). At the N-terminal end, Leu 201 of the JAZ1 peptide is embedded in a hydrophobic cavity presented by surface loops on top of COI1 (Fig. 3c). At the C-terminal end, Ala 204 of JAZ1 uses its short side chain to pack against the keto group of the ligand and Phe 89 of COI1 (Fig. 3c and Supplementary Fig. 4a). The same alanine residue of JAZ1 also donates a hydrogen bond through its backbone amide group to the keto moiety of the ligand emerging from the pocket (Fig. 3c). The middle region of the five-amino-acid sequence is secured to the COI1-jasmonate complex

bond and salt bridge networks are shown with yellow dashes. **f**, Top view of the JA-Ile pocket showing the  $F_o - F_c$  electron density, calculated before JA-Ile was built into the model (red mesh). The electron density of the pentenyl side chain of (3R,7S)-JA-Ile cannot accommodate the (3R,7R)-JA-Ile side chain, which is constrained by the chiral configuration at the C7 position. **g**, When bound to COI1, JA-Ile (yellow space fill) is solvent accessible at both the keto group (top) and carboxyl group (bottom).

through a hydrogen bond formed between the backbone carbonyl of Pro 202 in JAZ1 and the ligand-interacting COI1 residue Arg 496, which is critical for the hormone-dependent COI1-JAZ1 interaction (Supplementary Fig. 3). In agreement with its important role in forming the JA-Ile co-receptor, this short N-terminal region of the JAZ1 degron completely covers the opening of the ligand-binding pocket, conferring high-affinity binding to the hormone. The close interaction between the hormone and the co-receptor complex provides a plausible structural explanation for the favourable binding of the (3R,7S)-JA-Ile isomer, as the stereochemistry at the 7 position of (3R,7R)-JA-Ile may



**Figure 3 | The bi-partite JAZ1 degron peptide.** **a**, Top view of the complete JAZ1 degron peptide (orange) bound to COI1 (green) and JA-Ile (yellow). **b**, Side view and surface representation of the JAZ1 peptide, which acts as a clamp to lock JA-Ile in the pocket. **c**, Interactions of the N-terminal region of the JAZ1 degron with COI1 and JA-Ile. Hydrogen bonds are shown with yellow dashes. **d**, Structural role of the Arg 206 residue from the JAZ1 degron in coordinating the carboxyl group of JA-Ile with three basic residues of the COI1 ligand pocket floor. **e**, Top view of the amphipathic JAZ1 degron helix bound to COI1 with three hydrophobic residues of JAZ1 shown in stick representation (orange) and COI1 residues in coloured surface representation. **f**, Coronatine-induced interactions of wild-type and mutant COI1 with JAZ1 detected by a yeast two-hybrid assay. sdm, site-directed mutants. Blue colour indicates interaction.



place the aliphatic chain unfavourably close to nearby JAZ1 and COI1 residues (Supplementary Fig. 4a).

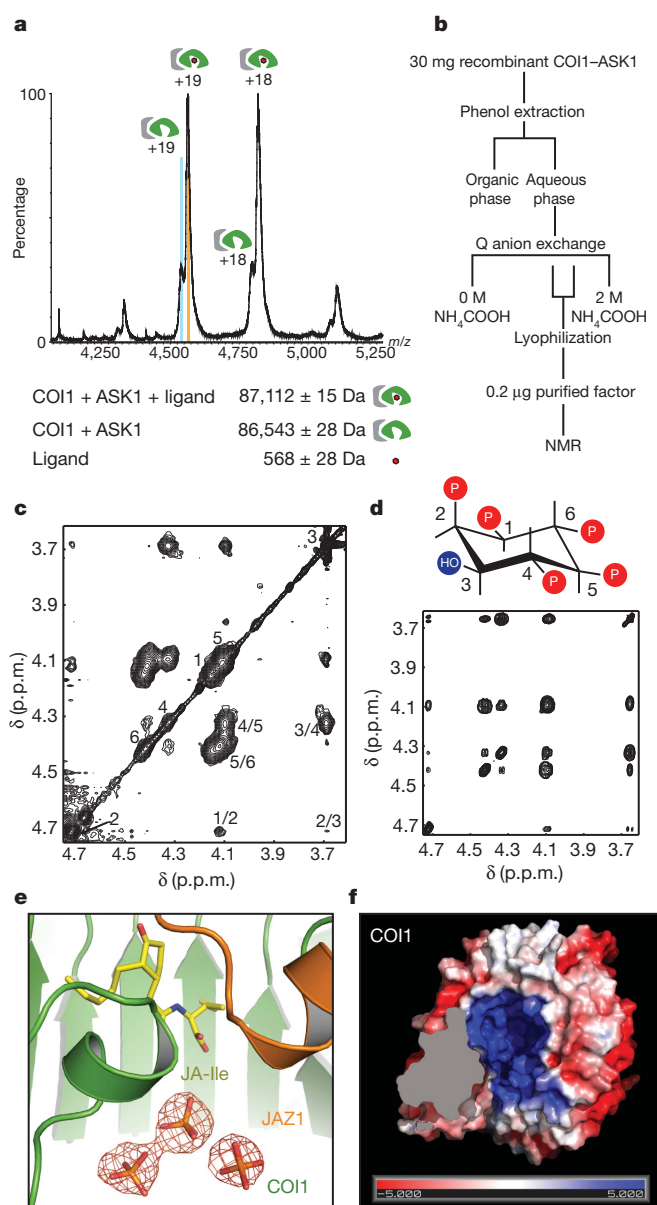
Within the JAZ1 degron, two conserved basic residues, Arg 205 and Arg 206, were previously shown to have an important role in hormone-induced COI1 binding<sup>17</sup>. In the structure, Arg 205 contributes to COI1 binding by directly interacting with loop-12, whereas Arg 206 points in the opposite direction and inserts deeply into the central tunnel of the COI1 solenoid. Approaching the bottom of the ligand-binding pocket, the guanidinium group of the Arg 206 side chain joins the three basic COI1 residues that form the pocket floor and interacts directly with the carboxyl group of the ligand (Fig. 3d). Thus, the N-terminal seven amino acids (ELPIARR) of the JAZ1 degron peptide act as a clamp that wraps the ligand-binding pocket from top to bottom, closing it completely (Fig. 3b).

The highly conserved C-terminal half of the JAZ1 degron forms an amphipathic  $\alpha$ -helix that strengthens the JAZ1–COI1 interaction by binding to the top surface of the COI1 LRR domain, adjacent to the ligand-binding site (Fig. 3a). With its N-terminal end directly packing against loop-2 of COI1, the Jas motif helix blocks the central tunnel of the COI1 LRR solenoid like a plug. The N-terminal half of the Jas motif helix is characterized by three hydrophobic residues—Leu 209, Phe 212 and Leu 213—which are aligned on the same side of the helix and form a hydrophobic interface with COI1 (Fig. 3e). By soaking the COI1–ASK1 crystals with coronatine and a sufficiently high concentration of JAZ1 degron peptide lacking the N-terminal ELPIA sequence, we were able to trap a complex formed by COI1, coronatine and the isolated Jas motif helix in the crystal (Supplementary Table 1). This indicates that the  $\alpha$ -helix may provide a low-affinity anchor for docking the JAZ protein on COI1. In support of this idea, single-amino-acid mutations at the complementary surface on COI1 readily disrupt hormone-induced COI1–JAZ1 interaction (Fig. 3f).

### Inositol pentakisphosphate as a cofactor of COI1

The crystal structure of TIR1 revealed an unexpected inositol hexakisphosphate ( $\text{InsP}_6$ ) molecule bound in the centre of the protein underneath the auxin-binding pocket<sup>20</sup>. The sequence homology between COI1 and TIR1 suggests that COI1 might also contain a similar small molecule. Before crystallization, we analysed the recombinant COI1–ASK1 complex by structural mass spectrometry. Nano-electrospray mass spectra of the intact COI1–ASK1 complex revealed two populations differing by a mass of  $\sim 568$  Da, indicating that a small molecule was indeed co-purified with the proteins (Fig. 4a and Supplementary Fig. 5). The mass-spectrometry-derived molecular mass of the unknown compound is different from the mass of  $\text{InsP}_6$  (651 Da) but matches that of an inositol pentakisphosphate ( $\text{InsP}_5$ ) molecule. Unfortunately, mass spectrometry analyses of either the native COI1–ASK1 complex or the denatured proteins were unable to achieve direct mass analysis of the small molecule.

To investigate the identity of the unknown compound, we first estimated that the molecule contains four or five phosphate groups by  $^{31}\text{P}$  nuclear magnetic resonance (NMR) of trypsin-digested COI1–ASK1 complex (data not shown). To identify unequivocally the unknown molecule, we set out to purify it away from the COI1–ASK1 complex in a quantity sufficient for  $^1\text{H}$  NMR analysis. The high phosphate content of the molecule allowed us to trace it through a multi-step purification procedure (Fig. 4b). After isolation of 150 nmol of the purified small molecule, we acquired a series of one-dimensional and two-dimensional NMR data, including a highly informative homonuclear total correlation (TOCSY) spectrum. The observed chemical shifts and TOCSY cross-peak patterns are clearly characteristic of inositol phosphates (Fig. 4c). A comparison with previously reported NMR spectra of various inositol phosphates established that the unknown compound is either D- or L-inositol-1,2,4,5,6-pentakisphosphate ( $\text{Ins}(1,2,4,5,6)\text{P}_5$ ; Fig. 4c)<sup>21</sup>. This conclusion was further supported by the TOCSY spectrum of synthetic  $\text{Ins}(1,2,4,5,6)\text{P}_5$  (Fig. 4d) and the subsequently acquired negative-ion electrospray



**Figure 4 | Identification of an inositol pentakisphosphate cofactor in COI1.**

**a**, Nano-electrospray mass spectrometry of the intact COI1–ASK1 complex. Low-intensity charge series corresponds in mass to the cofactor-free COI1–ASK1 complex. High-intensity charge series corresponds to the cofactor-bound COI1–ASK1 complex. **b**, Optimized cofactor purification scheme. **c**, Proton TOCSY spectrum of the purified cofactor. Numbers along the diagonal indicate the positions of the six protons of  $\text{Ins}(1,2,4,5,6)\text{P}_5$ . The cross-peaks corresponding to direct couplings are labelled. Other cross-peaks correspond to relayed connectivities. p.p.m., parts per million. **d**, TOCSY spectrum of a synthetic  $\text{Ins}(1,2,4,5,6)\text{P}_5$  as a standard. **e**, Islands of positive  $F_o - F_c$  electron density (red mesh) below the hormone-binding pockets, which probably belong to inorganic phosphate molecules from the crystallization solutions that displace  $\text{InsP}_5$  from the  $\text{InsP}_5$ -binding site. **f**, Bottom view of a surface electrostatic potential representation of COI1 from positive (blue) to negative (red).

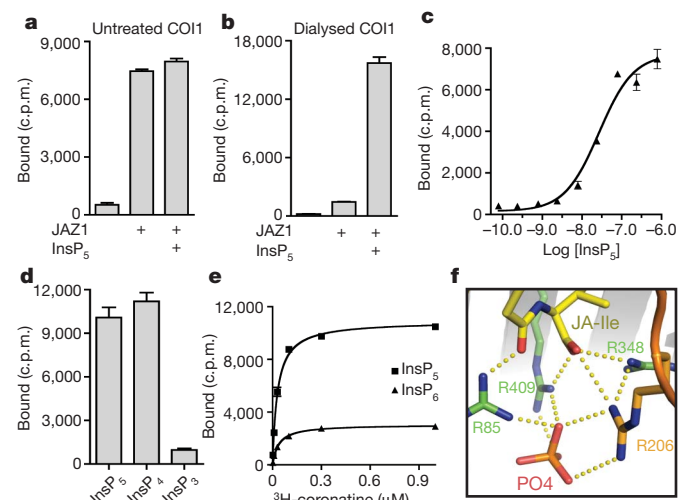
ionization mass spectrometry spectrum of the compound (Supplementary Fig. 6).

Consistent with the binding of a small molecule cofactor, the crystal structure of COI1 showed strong unexplained electron densities clustered in the middle of the COI1 LRR domain. Like  $\text{InsP}_6$  in TIR1, these extra densities in COI1 are located directly adjacent to the bottom of the ligand-binding pocket of the jasmonate co-receptor, interacting with multiple positively charged COI1 residues (Fig. 4e). Unexpectedly, these islands of electron density cannot be explained by an  $\text{Ins}(1,2,4,5,6)\text{P}_5$  molecule.

Instead, their intensity, overall symmetry and poor connectivity indicate that they belong to multiple free phosphate molecules. Because a high concentration of ammonium phosphate was used as the major precipitant for crystallizing the jasmonate co-receptor, we postulate that the  $\text{InsP}_5$  molecule that co-purified with COI1 was later displaced by phosphate molecules in the crystallization drops. In support of this scenario, the concave surface of the COI1 solenoid fold surrounding the phosphates is highly basic and decorated with residues conserved in plant COI1 orthologues, indicating a functionally important surface area (Fig. 4f and Supplementary Figs 2 and 7).

### $\text{InsP}_5$ potentiates jasmonate perception by COI1–JAZ1

The highly selective co-purification of two different inositol phosphates,  $\text{InsP}_5$  and  $\text{InsP}_6$ , with two homologous plant hormone receptors, COI1 and TIR1, implies that the proper function of the two F-box proteins might require the binding of specific inositol phosphates. To assess the functional role of  $\text{Ins}(1,2,4,5,6)\text{P}_5$  in the COI1–JAZ1 co-receptor, we took advantage of our crystallographic observation and developed a protocol to strip the co-purified  $\text{InsP}_5$  from COI1 without denaturing the protein. The resulting COI1–ASK1 complex was then tested in a ligand-binding-based reconstitution assay. As shown in Fig. 5a, untreated COI1 formed a high-affinity jasmonate co-receptor with JAZ1. Addition of exogenous  $\text{Ins}(1,2,4,5,6)\text{P}_5$  did not significantly change its activity. In contrast, the dialysed COI1 sample completely lacked ligand binding by itself and showed only trace activity in the presence of JAZ1. Supplementation with either synthetic  $\text{Ins}(1,2,4,5,6)\text{P}_5$  (Fig. 5b) or the purified and NMR-analysed  $\text{InsP}_5$  sample (data not shown) rescued the interaction in a dose-dependent manner and with a half-maximum effective concentration ( $\text{EC}_{50}$ ) of 27 nM (Fig. 5c). From this reconstitution result, we conclude that  $\text{Ins}(1,2,4,5,6)\text{P}_5$  binding is crucial for the jasmonate co-receptor to perceive the hormone with high sensitivity.



**Figure 5 | Inositol phosphate is an essential component of the COI1–JAZ co-receptor.** **a**, Binding of  $^3\text{H}$ -coronatine at 100 nM to a complex of COI1 and JAZ1, with the addition of 1  $\mu\text{M}$  synthetic  $\text{Ins}(1,2,4,5,6)\text{P}_5$  ( $\text{InsP}_5$ ). **b**, With extensive dialysis to remove the co-purified  $\text{InsP}_5$  cofactor, 100 nM  $^3\text{H}$ -coronatine no longer binds dialysed COI1 in the presence of JAZ1. Synthetic  $\text{Ins}(1,2,4,5,6)\text{P}_5$  rescues binding. **c**,  $\text{Ins}(1,2,4,5,6)\text{P}_5$  rescues the binding of 100 nM  $^3\text{H}$ -coronatine to dialysed COI1–ASK1 in the presence of JAZ1 with an  $\text{EC}_{50}$  of  $27 \pm 12$  nM. **d**, Binding assays performed with 100 nM  $^3\text{H}$ -coronatine, dialysed COI1 and 1  $\mu\text{M}$  synthetic  $\text{Ins}(1,2,4,5,6)\text{P}_5$  ( $\text{InsP}_5$ ),  $\text{Ins}(1,4,5,6)\text{P}_4$  ( $\text{InsP}_4$ ), or  $\text{Ins}(1,4,5)\text{P}_3$  ( $\text{InsP}_3$ ). **e**, Saturation binding of  $^3\text{H}$ -coronatine to dialysed COI1 in the presence of 1  $\mu\text{M}$  of  $\text{Ins}(1,2,4,5,6)\text{P}_5$  ( $\text{InsP}_5$ ) and  $\text{Ins}(1,2,3,4,5,6)\text{P}_6$  ( $\text{InsP}_6$ ) at a  $K_d$  of  $30 \pm 5$  nM and  $37 \pm 8$  nM, respectively. All results are the mean  $\pm$  s.e. of up to three experiments performed in duplicate. **f**, A phosphate-binding site in the complex structure reveals an interwoven hydrogen bond network that may explain the mechanism by which the  $\text{InsP}$  cofactor potentiates the jasmonate co-receptor.

Although further effort is needed to reveal how  $\text{InsP}_5$  binds to COI1, a close examination of the phosphate molecules in the available COI1 structure indicates a mechanism by which the inositol phosphate molecule may modulate the activity of the jasmonate co-receptor. Among four COI1-bound phosphates, one stands out by binding at a critical position in the jasmonate co-receptor. This phosphate molecule interacts simultaneously with four basic residues at the bottom of the ligand-binding pocket, namely Arg 206 in the JAZ1 degron and the three COI1 arginine residues that form the floor of the pocket. As a result, a tetragonal bipyramidal interaction network is formed among four molecules at the core of the jasmonate co-receptor assembly. The four arginines from COI1 and JAZ1 sit at the four corners of the central plane, interacting with the hormone above and the phosphate below (Fig. 5f). As the free phosphate molecule probably mimics the action of a phosphate group on  $\text{InsP}_5$ , this four-molecule junction, together with additional phosphate–COI1 interactions seen in the crystal, conceivably represent the structural basis for  $\text{InsP}_5$  potentiation of the jasmonate co-receptor. Consistent with this interpretation, coronatine-induced formation of a COI1–JAZ1 complex was readily abolished by mutation of select COI1 residues adjacent to the phosphates, but not in contact with the hormone (Supplementary Fig. 8).

We used the reconstitution assay to investigate further the specificity of jasmonate co-receptor regulation by inositol phosphates (Fig. 5d). Notably, inositol-1,4,5,6-tetrakisphosphate supports the activity of the COI1–JAZ1 co-receptor, whereas the second messenger signalling molecule inositol-1,4,5-trisphosphate does not. Addition of a phosphate to  $\text{InsP}_5$ , which gives rise to  $\text{InsP}_6$ , is also not favourable for activity. Although saturation binding of  $^3\text{H}$ -coronatine is stimulated by both  $\text{Ins}(1,2,4,5,6)\text{P}_5$  and  $\text{InsP}_6$  with similar  $K_d$  values (30 nM and 37 nM, respectively), the two inositol phosphates yield markedly different  $B_{\text{max}}$  values for coronatine binding, indicating that  $\text{InsP}_6$  is significantly less efficacious in activating the co-receptor despite having equal affinity as  $\text{Ins}(1,2,4,5,6)\text{P}_5$  (Fig. 5e). Functional selectivity of COI1 for the inositol phosphate cofactor is consistent with the conservation of the putative inositol-phosphate-binding site, which is distinct in amino acid sequence from the  $\text{InsP}_6$ -binding site in TIR1<sup>20</sup> (Supplementary Fig. 2).

### Discussion

Our structural and pharmacological analyses reveal not only the essential components of the receptor system but also the detailed mechanism by which these components cooperatively assemble and recognize the hormonal signal through a network of interactions. Our data identify the true jasmonate receptor as a three-molecule co-receptor complex, consisting of COI1, JAZ degron and inositol pentakisphosphate, all of which are indispensable for high-affinity hormone binding. Our analyses also define the JAZ degron boundaries as a unique bi-partite sequence that binds COI1 and directly participates in hormone recognition. Unexpectedly, the N-terminal clamp region of the JAZ1 degron that is critical for hormone binding is diverse among JAZ proteins. This variable sequence might create a family of COI1–JAZ co-receptors that respond differentially to the hormone.

The crystal structure of the COI1–JAZ1 co-receptor in complex with JA-Ile revealed a markedly different binding mode of the hormone as predicted by computational modelling<sup>12</sup>. Although COI1 shares high sequence homology with TIR1, subtle structural differences and the integration of two additional factors critical for ligand binding give rise to a hormone-binding pocket in COI1 that is challenging to model. For the same reason, the structural nature of the ligand-free form of the F-box protein cannot be modelled with accuracy. The direct interactions of the hormone with both COI1 and the JAZ protein as observed in the crystal nonetheless support a molecular glue mechanism previously proposed for the auxin system<sup>20</sup>.

Discovery of the inositol pentakisphosphate cofactor of COI1 has important implications for the role of inositol phosphates in plant hormone signalling. COI1 co-purifies with a single isoform of



InsP<sub>5</sub>, Ins(1,2,4,5,6)P<sub>5</sub>, indicating selectivity at the receptor level. However, both inositol-1,2,4,5,6-pentakisphosphate and inositol-1,4,5,6-tetrakisphosphate support high-affinity hormone binding in our reconstitution assays, leaving the identity of the physiologically relevant form of inositol phosphate an open question.

Finally, our study is the latest in a series of receptor structures for plant hormones, including auxin<sup>20</sup>, gibberellin<sup>22,23</sup> and abscisic acid<sup>24–28</sup>. Despite different structural mechanisms, a common theme of hormone-mediated protein interactions emerges as a unique strategy favoured by plant systems throughout evolution.

## METHODS SUMMARY

The Methods provides detailed information about all experimental procedures, including: (1) description of protein preparation, purification and mutagenesis; (2) description of protein crystallization, data collection and structure determination; (3) details for conducting *in vitro* radioligand binding assay; (4) details for conducting yeast two-hybrid assay; (5) description of inositol phosphate purification scheme; (6) details for conducting *in vitro* inositol phosphate reconstitution assays; (7) description of structural mass spectrometry analysis of the intact protein complex; (8) description of NMR analysis of the inositol phosphate; and (9) description of mass spectrometry analysis of the inositol phosphate.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 30 June; accepted 19 August 2010.

Published online 6 October 2010.

- Browse, J. Jasmonate passes muster: a receptor and targets for the defense hormone. *Annu. Rev. Plant Biol.* **60**, 183–205 (2009).
- Feys, B., Benedetti, C. E., Penfold, C. N. & Turner, J. G. *Arabidopsis* mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. *Plant Cell* **6**, 751–759 (1994).
- Staswick, P. E. & Tiryaki, I. The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis*. *Plant Cell* **16**, 2117–2127 (2004).
- Fonseca, S. *et al.* (+)-7-iso-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. *Nature Chem. Biol.* **5**, 344–350 (2009).
- Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M. & Turner, J. G. COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280**, 1091–1094 (1998).
- Chini, A. *et al.* The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448**, 666–671 (2007).
- Thines, B. *et al.* JAZ repressor proteins are targets of the SCF<sup>COI1</sup> complex during jasmonate signalling. *Nature* **448**, 661–665 (2007).
- Yan, Y. *et al.* A downstream mediator in the growth repression limb of the jasmonate pathway. *Plant Cell* **19**, 2470–2483 (2007).
- Lorenzo, O., Chico, J. M., Sanchez-Serrano, J. J. & Solano, R. JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in *Arabidopsis*. *Plant Cell* **16**, 1938–1950 (2004).
- Kepinski, S. & Leyser, O. The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* **435**, 446–451 (2005).
- Dharmasiri, N., Dharmasiri, S. & Estelle, M. The F-box protein TIR1 is an auxin receptor. *Nature* **435**, 441–445 (2005).
- Yan, J. *et al.* The *Arabidopsis* CORONATINE INSENSITIVE1 protein is a jasmonate receptor. *Plant Cell* **21**, 2220–2236 (2009).
- Katsir, L., Schilmiller, A. L., Staswick, P. E., He, S. Y. & Howe, G. A. COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. *Proc. Natl Acad. Sci. USA* **105**, 7100–7105 (2008).
- Suza, W. P. & Staswick, P. E. The role of JAR1 in jasmonoyl-L-isoleucine production during *Arabidopsis* wound response. *Planta* **227**, 1221–1232 (2008).
- Koo, A. J., Gao, X., Jones, A. D. & Howe, G. A. A rapid wound signal activates the systemic synthesis of bioactive jasmonates in *Arabidopsis*. *Plant J.* **59**, 974–986 (2009).
- Chung, H. S. & Howe, G. A. A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in *Arabidopsis*. *Plant Cell* **21**, 131–145 (2009).
- Melotto, M. *et al.* A critical role of two positively charged amino acids in the Jas motif of *Arabidopsis* JAZ proteins in mediating coronatine- and jasmonoyl isoleucine-dependent interactions with the COI1 F-box protein. *Plant J.* **55**, 979–988 (2008).
- Grunewald, W. *et al.* Expression of the *Arabidopsis* jasmonate signalling repressor JAZ1/TIFY10A is stimulated by auxin. *EMBO Rep.* **10**, 923–928 (2009).
- Chung, H. S. *et al.* Alternative splicing expands the repertoire of dominant JAZ repressors of jasmonate signaling. *Plant J.* **63**, 613–622 (2010).
- Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640–645 (2007).
- Stephens, L. R. *et al.* myo-inositol pentakisphosphates. Structure, biological occurrence and phosphorylation to myo-inositol hexakisphosphate. *Biochem. J.* **275**, 485–499 (1991).
- Shimada, A. *et al.* Structural basis for gibberellin recognition by its receptor GID1. *Nature* **456**, 520–523 (2008).
- Murase, K., Hirano, Y., Sun, T. P. & Hakoshima, T. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **456**, 459–463 (2008).
- Santiago, J. *et al.* The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature* **462**, 665–668 (2009).
- Melcher, K. *et al.* A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. *Nature* **462**, 602–608 (2009).
- Miyazono, K. *et al.* Structural basis of abscisic acid signalling. *Nature* **462**, 609–614 (2009).
- Nishimura, N. *et al.* Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. *Science* **326**, 1373–1379 (2009).
- Yin, P. *et al.* Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nature Struct. Mol. Biol.* **16**, 1230–1236 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the beamline staff of the Advanced Light Source at the University of California at Berkeley and the Advanced Photon Source at Argonne National Laboratory for help with data collection. We also thank P. Rajagopal and R. Klevit for <sup>31</sup>P NMR analysis, M. Sadilek for mass spectrometry analysis, L. Katsir, C. Hague and J. Lyssand for radioligand binding assay assistance, and C. Mecey and M. Melotto for initial cloning of coi1(sdm) mutants. We also thank members of the Zheng laboratory and W. Xu laboratory and R. Gardner for discussion and help. This work is supported by the Howard Hughes Medical Institute and grants from the National Institutes of Health (R01 CA107134 to N.Z., T32 GM07270 to L.B.S., R01GM57795 to G.A.H., R01AI068718 to S.Y.H.), National Science Foundation (0929100 to N.Z.), US Department of Energy (DE-FG02-99ER20323 to J.B. and DE-FG02-91ER20021 to G.A.H. and S.Y.H.), Michigan State University Plant Science Fellowship (J.W.), the Welch Foundation (I-1304 to J.R.) and the European Research Council (ERC) under the European Community's Seventh Framework Program (FP7/2007-2013)/ ERC Grant agreement no. 239679 (G.B.-N. and M.S.).

**Author Contributions** L.B.S., G.A.H. and N.Z. conceived and L.B.S. conducted radioligand binding and additional functional experiments. X.T., H.M. and L.B.S. purified the COI1-ASK1 complex and conducted crystallographic experiments. X.T. crystallized and determined the structures of the COI1-ASK1-JAZ1 hormone complexes. L.B.S., X.T. and N.Z. analysed crystallographic data. J.W. and S.Y.H. conceived and J.W. conducted yeast two-hybrid experiments. G.B.-N. and M.S. conducted and interpreted the structural mass spectrometry experiments. L.B.S., H.M., T.R.H., F.-F.H., J.R. and N.Z. conceived and conducted experiments for inositol phosphate purification and identification. Y.K. synthesized jasmonate stereoisomers. L.B.S. and N.Z. wrote the manuscript with comments from all authors.

**Author Information** Structural coordinates and structural factors have been deposited in the Protein Data Bank under accession numbers 3OGK, 3OGL and 3OGM. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to N.Z. (nzheng@u.washington.edu).

## METHODS

**Protein preparation.** The full-length *Arabidopsis thaliana* COI1 and ASK1 were co-expressed as a glutathione S-transferase (GST) fusion protein and an untagged protein, respectively, in *Hi5* suspension insect cells. The COI1–ASK1 complex was isolated from the soluble cell lysate by glutathione affinity chromatography. After on-column tag cleavage by tobacco etch virus protease, the complex was further purified by anion exchange and gel filtration chromatography and concentrated by ultrafiltration to  $12\text{--}18\text{ mg ml}^{-1}$ . Full-length JAZ substrate proteins were expressed as 6×His-fusion proteins in *Escherichia coli* and purified on Ni-NTA resin with subsequent dialysis into 20 mM Tris-HCl, pH 8.0, 200 mM NaCl and 10% glycerol. For truncation mutants, a stop codon was introduced in JAZ1 proteins using the Quick-Change II site-directed mutagenesis kit (Stratagene). Synthetic JAZ degron peptides were prepared by United Biochemical Research, Inc. JAZ degron fusion peptides were prepared with N-terminal 6×-His tag and C-terminal GST fusion tag and expressed in *E. coli*. The protein was isolated by glutathione affinity resin for pull-down assay with untagged COI1–ASK1 complex.

**Site-directed mutagenesis.** Individual amino acid residues in the LRR domain of COI1 proteins were mutated to alanine using the Quick-Change II site-directed mutagenesis kit (Stratagene). Mutant proteins were co-expressed with JAZ1 (JAZ1:pB42AD) in yeast to detect protein–protein interactions.

**Crystallization, data collection and structure determination.** The crystals of the COI1–ASK1–JAZ1 peptide complexes bound to either coronatine or JA-Ile were grown at 4 °C by the hanging-drop vapour diffusion method with 1.5 µl protein complex samples containing COI1–ASK1, JAZ1 peptide and hormone compound at 1:1:1 molar ratio mixed with an equal volume of reservoir solution containing 100 mM BTP, 1.7–1.9 M ammonium phosphate, 100 mM NaCl, pH 7.0. Diffraction quality crystals were obtained by the micro-seeding method at 4 °C. The crystals all contain eight copies of the complex in the asymmetric unit. The data sets were collected at the BL8.2.1 beamline at the Advanced Light Source in Lawrence Berkeley National Laboratory as well as the GM/CA-CAT 23 ID-B beamline at the Advanced Photon Source in Argonne National Laboratory using crystals flash-frozen in the crystallization buffers supplemented with 15–20% ethylene glycol at –170 °C. Reflection data were indexed, integrated and scaled with the HKL2000 package<sup>29</sup>. All crystal structures were solved by molecular replacement using the program Phaser<sup>30</sup> and the TIR1–ASK1 structure as search model. The structural models were manually built in the program O<sup>31</sup> and refined using CNS<sup>32</sup> and PHENIX<sup>30</sup>. All final models have 96–98% of residues in the favoured region and 0% in disallowed region of the Ramachandran plot.

**Hormone and inositol phosphate reagents.** <sup>3</sup>H-coronatine was synthesized by Amersham<sup>13</sup>. Coronatine was purchased from Sigma; JA-Ile conjugates were chemically synthesized as previously described<sup>33</sup>. Synthetic inositol phosphates were purchased as sodium salts from Cayman Chemicals. InsP<sub>6</sub> was purchased from Sigma.

**Radioligand binding assay.** Radioligand binding was assayed on purified proteins, with 2 µg COI1–ASK1 complex and JAZ proteins at a 1:3 molar ratio, and/or 10 µM synthetic peptides. Reactions were prepared in 100 µl final volume and in a binding buffer containing 20 mM Tris-HCl, 200 mM NaCl and 10% glycerol. Saturation binding experiments were conducted with serial dilutions of <sup>3</sup>H-coronatine in binding buffer. Nonspecific binding was determined in the presence of 300 µM coronatine. Competition binding experiments were conducted with serial dilutions of JA-Ile in the presence of 100 nM <sup>3</sup>H-coronatine with nonspecific binding determined in the presence of 300 µM coronatine. Total binding was determined in the presence of vehicle only. Two-point binding experiments were performed in the presence of 100 nM or 300 nM <sup>3</sup>H-coronatine with nonspecific binding determined in the presence of 300 µM coronatine. Following incubation with mixing at 4 °C, all samples were collected with a cell harvester (Brandel, Gaithersburg, MD) on polyethyleneimine (Sigma)-treated filters. Samples were incubated in liquid scintillation fluid for >1 h before counting with a Packard Tri-Carb 2200 CA liquid scintillation analyser (Packard Instrument Co.). Saturation binding experiments were analysed by nonlinear regression, competition binding experiments by nonlinear regression with K<sub>i</sub> calculation as per the method of ref. 34, and concentration–response data by sigmoidal dose–response curve fitting, all using GraphPad Prism version 4.00 for MacOSX.

**Yeast two-hybrid assay.** The coding sequences (CDS) of the *Arabidopsis thaliana* gene *COI1* (At2g39940) and *coi1* site-directed mutants were cloned into the yeast two-hybrid bait vector pGILDA (Clontech) using XmaI and XhoI restriction enzyme recognition sequences previously added to the 5' and 3' end of the *COI1* CDS, respectively, creating DNA-binding domain (LexA–COI1 and LexA–coi1) protein fusions. The CDS of *Arabidopsis thaliana* JAZ1 gene (At1g19180) was cloned into the yeast two-hybrid prey vector pB42AD (Clontech) creating a transcriptional activation domain (AD–JAZ1) fusion protein. Individual wild-type and mutant *COI1* constructs were co-transformed with JAZ1 constructs into *Saccharomyces cerevisiae* strain EGY48 (p8opLacZ) using the frozen-EZ yeast transformation II

kit (Zymo Research). Transformants were selected on SD-glucose medium (BD Biosciences) supplemented with –Ura/–Trp/–His drop-out solution (BD Biosciences). To detect the interaction between COI1 and JAZ1, transformants that had been selected in SD-Glu medium were re-suspended in sterile water. Ten microlitres of each suspension was spotted onto inducing media (SD-Galactose/Raffinose –UWH; BD Biosciences) supplemented with 80 µg ml<sup>–1</sup> X-Gal and 50 µM coronatine (Sigma). Yeast two-hybrid assay plates were incubated in the dark at 20 °C and photographed 7 days later. Induced yeast cells were analysed for COI1 and JAZ1 expression levels by western blotting using epitope-specific antibodies (data not shown).

**Inositol phosphate purification.** Phenol was melted at 68 °C and equilibrated with equal parts 0.5 M Tris-HCl, pH 8.0 until a pH of 7.8 was reached. The equilibrated phenol was then topped with 0.1 volume 100 mM Tris-HCl, pH 8.0 and stored at 4 °C. For extraction, 30–40 mg of 1 mg ml<sup>–1</sup> COI1–ASK1 protein was mixed in small batches with equal parts equilibrated phenol at room temperature. The samples were inverted and incubated for 30 min until phase separation occurred. With 30 s vortexing, the samples were incubated at room temperature for 30 min and spun at 15,000 r.p.m. for 5 min. The aqueous phase was removed as a primary extraction. Equal parts of a solution containing 25 mM Tris-HCl, pH 8.0 was added to the phenol and collected as above as a secondary extraction. The primary and secondary extractions were then combined and diluted 10× in 25 mM Tris-HCl, pH 8.0, then further purified by gravity flow on Q sepharose high-performance anion exchange resin (GE Healthcare). Following column wash with 10× column volumes of 0.1 N formic acid, stepwise elution was performed with 2× column volumes of 0.1 N formic acid (Thermo Scientific) with increasing concentrations of ammonium formate (Sigma), from 0 to 2 M.

Fractions were analysed for phosphate content by the wet-ashing method with perchloric acid in Pyrex culture tubes (13 × 100 mm). Typically, samples of 50–100 µl were ashed with 100–200 µl 70% perchloric acid (purified by redistillation, Sigma). Ashing was performed by heating the sample over a Bunsen-type burner with continuous shaking to prevent bumping. When the sample stopped emitting white smoke, the reaction was considered complete and then heated to dryness. 500 µl of distilled water was added to the room temperature tubes and vortexed. 100 µl samples containing up to 10 nmol inorganic phosphate were assayed for phosphate by a modification of a published procedure<sup>35</sup>. A total of 125 µl of acid molybdate colour reagent was added and the samples were incubated and covered at room temperature for 12–14 h (overnight) for full colour development (total volume 225 µl). Plates were read at 650 nm and unknowns were determined from the linear regression of the standard curve (0–10 nmol NaH<sub>2</sub>PO<sub>4</sub> per well). All assays were done in triplicate. Final fractions containing phosphate were combined and lyophilized repeatedly to remove residual ammonium formate.

**Inositol phosphate reconstitution assays.** COI1–ASK1 complex was separated from pre-bound inositol phosphate by dialysis. Briefly, proteins were mixed with 10% glycerol and incubated in 2 M ammonium phosphate, 100 mM Bis-Tris propane pH 7.0, 200 mM NaCl, 10% glycerol, at 4 °C for >24 h with a minimum of 3× buffer changes at 100× sample volume. Samples were then transferred to 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10% glycerol, at 4 °C for >24 h with a minimum of three buffer changes at 100× sample volume. Inositol phosphate rescue experiments were conducted according to the radioligand binding assays described above in the presence of 300 nM <sup>3</sup>H-coronatine with nonspecific binding determined in the presence of 300 µM coronatine.

**Structural mass spectrometry analysis of the intact protein complex.** Nano-electrospray ionization mass spectrometry (MS) and tandem MS (MS/MS) experiments were performed on a Synapt HDMS instrument. Before MS analysis, 50 µl of a 16 mg ml<sup>–1</sup> solution of COI1–ASK1 in 20 mM Tris-HCl pH 8, 0.2 M NaCl and 5 mM DTT, was buffer-exchanged twice into 0.5 M ammonium acetate solution by using Bio-Rad Biospin columns. To improve desolvation during ionization, samples were diluted 1:4 in 0.5 M ammonium acetate and isopropanol was added to a final concentration of 5%. Typically an aliquot of 2 µl solution was loaded for sampling via nano-ESI capillaries which were prepared in-house from borosilicate glass tubes as described previously<sup>36</sup>. The conditions within the mass spectrometer were adjusted to preserve non-covalent interactions. The following experimental parameters were used: capillary voltage up to 1.26 kV, sampling cone voltage 150 V and extraction cone voltage 6 V, MCP 1590. For tandem MS experiments peaks centred at *m/z* 4,564 and 4,588 were selected in the quadrupole and collision energy up to 65 V was used. Argon was used as a collision gas at maximum pressure. All spectra were calibrated externally by using a solution of caesium iodide (100 mg ml<sup>–1</sup>). Spectra are shown with minimal smoothing and without background subtraction.

**Nuclear magnetic resonance (NMR) analysis.** NMR spectra were acquired on a Varian INOVA600 spectrometer equipped with a cold probe using 200 µM samples of purified compound X or synthetic inositol-1,2,4,5,6-pentakisphosphate (Cayman

Chemical) dissolved in D<sub>2</sub>O. TOCSY spectra were acquired with mixing times of 35 or 50 ms, processed with NMRPipe<sup>37</sup> and visualized with NMRView<sup>38</sup>.

#### Mass spectrometry analysis of inositol phosphate purified from COI1-ASK1.

MS experiments were conducted on a Finnigan LTQ linear ion-trap mass spectrometer (ITMS) with Xcalibur operating system. Methanol was continuously infused (10 µl min<sup>-1</sup>) to the ESI source, where the skimmer was set at ground potential, the electrospray needle was set at 4.5 kV, and the temperature of the heated capillary was 275 °C. The sample was diluted with equal volume of 2% ammonia in methanol and 10 µl was flow injected. The automatic gain control of the ion trap was set at  $2 \times 10^4$ , with a maximum injection time of 50 ms. Helium was used as the buffer and collision gas at a pressure of  $1 \times 10^{-3}$  mbar (0.75 mTorr). The MS<sup>n</sup> ( $n = 2, 3, 4, 5$ ) experiments were carried out with an optimized relative collision energy ranging from 12% to 16% with an activation  $q$  value at 0.25. The activation time was set at 30–60 ms. The mass spectra were acquired in the profile mode and were accumulated for 3–5 min for MS<sup>n</sup> spectra. The mass resolution of the instrument was tuned to 0.6 Da at half peak height.

29. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic, 1997).
30. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
31. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
32. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
33. Ogawa, N. & Kobayashi, Y. Strategy for synthesis of the isoleucine conjugate of epigallocatechin gallate. *Tetrahedr. Lett.* **49**, 7124–7127 (2008).
34. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant ( $K_i$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).
35. Sadzadeh, S. M., Vincenzi, F. F. & Hinds, T. R. Simultaneous measurement of multiple membrane ATPases in microtiter plates. *J. Pharmacol. Toxicol. Methods* **30**, 103–110 (1993).
36. Nettleton, E. J. *et al.* Protein subunit interactions and structural integrity of amyloidogenic transthyretins: evidence from electrospray mass spectrometry. *J. Mol. Biol.* **281**, 553–564 (1998).
37. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
38. Johnson, B. A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352 (2004).



# An unprecedented nucleic acid capture mechanism for excision of DNA damage

Emily H. Robinson<sup>1</sup>, A. S. Prakasha Gowda<sup>2</sup>, Thomas E. Spratt<sup>2</sup>, Barry Gold<sup>3</sup> & Brandt F. Eichman<sup>1</sup>

DNA glycosylases that remove alkylated and deaminated purine nucleobases are essential DNA repair enzymes that protect the genome, and at the same time confound cancer alkylation therapy, by excising cytotoxic N3-methyladenine bases formed by DNA-targeting anticancer compounds. The basis for glycosylase specificity towards N3- and N7-alkylpurines is believed to result from intrinsic instability of the modified bases and not from direct enzyme functional group chemistry. Here we present crystal structures of the recently discovered *Bacillus cereus* AlkD glycosylase in complex with DNAs containing alkylated, mismatched and abasic nucleotides. Unlike other glycosylases, AlkD captures the extrahelical lesion in a solvent-exposed orientation, providing an illustration for how hydrolysis of N3- and N7-alkylated bases may be facilitated by increased lifetime out of the DNA helix. The structures and supporting biochemical analysis of base flipping and catalysis reveal how the HEAT repeats of AlkD distort the DNA backbone to detect non-Watson-Crick base pairs without duplex intercalation.

Alkylation of DNA by endogenous methyl donors, environmental toxins and chemotherapeutic agents produces a diverse spectrum of cytotoxic and mutagenic lesions, including N3-methyladenine (3mA), N7-methylguanine (7mG) and 1,N<sup>6</sup>-ethenoadenine (εA), that threaten the survival of all organisms<sup>1–5</sup>. 3mA is highly toxic owing to its inhibition of DNA polymerases during replication<sup>6,7</sup>, and production of such lesions is the rationale behind the use of alkylating agents in chemotherapy. N7-substituted guanines are the most prevalent alkylation lesions and display a wide range of toxic and mutagenic biological properties<sup>8</sup>. By virtue of their positive charges at physiological pH, 3mA and 7mG are especially susceptible to spontaneous depurination, which generates abasic sites in DNA that can ultimately lead to single- and double-strand breaks.

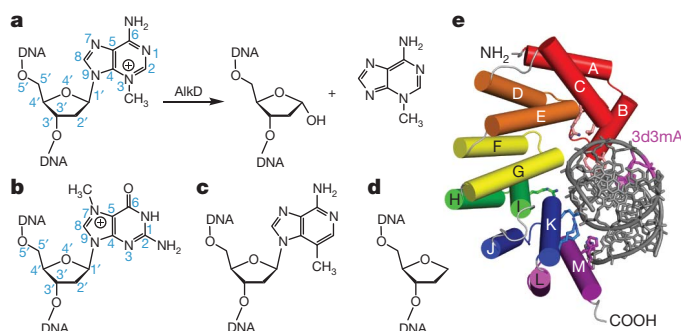
DNA glycosylases initiate base excision repair of N3- and N7-methylpurines from the genome by catalysing hydrolysis of the N-glycosidic bond (Fig. 1a, b). Despite their structural diversity, all

DNA glycosylases studied so far use a common base-flipping mechanism to access damaged DNA and orient the substrate for catalysis by rotating the target nucleotide 180° around the phosphoribose backbone into a complementarily shaped active site pocket<sup>9,10</sup>. The resulting distortion to the DNA is stabilized by an intercalating side-chain ‘plug’ that fills the void created by the extrahelical nucleotide. Glycosylases typically excise their target nucleobases by using a carboxylate side chain as a general base to activate a water nucleophile or to stabilize the carbocation transition state during base dissociation<sup>11</sup>. Mutation of this residue, however, does not abolish catalytic activity in all cases, leading to a model in which conformational strain in the DNA arising from extensive binding energy helps to drive the reaction forward<sup>12,13</sup>. The lack of a residue capable of performing general base catalysis in 3mA-specific DNA glycosylases (for example, *Escherichia coli* TAG)<sup>14–16</sup> is consistent with the idea that excision of positively charged 3mA and 7mG does not require the same level of catalytic assistance as more stable ethenoadducts, although direct evidence for this has not been reported.

AlkC and AlkD proteins, recently discovered in *Bacillus cereus* and subsequently identified in all three kingdoms of life (Supplementary Fig. 1), have emerged as a unique DNA glycosylase superfamily specific for N3- and N7-alkylpurines<sup>17,18</sup>. AlkD accelerates the rate of 7mG hydrolysis from DNA 100-fold over the spontaneous rate of 7mG depurination<sup>19</sup>, prompting us to investigate the mechanism by which AlkD excises destabilized alkylated bases. Here we present crystal structures of *B. cereus* AlkD in complex with DNA damage resembling the substrate and product of the glycosylase reaction. These structures, together with supporting biochemistry of base-flipping and 7mG depurination activities, demonstrate how AlkD uses an unprecedented strategy to trap non-canonical base pairs that allows for specific hydrolysis of destabilized N-glycosidic bonds without direct chemical attack from the enzyme.

## A new architecture for binding nucleic acids

We previously determined the crystal structure of *B. cereus* AlkD and identified residues important for DNA binding and catalysis<sup>19</sup>. AlkD



**Figure 1 | Base excision repair of alkylated DNA by AlkD.** **a**, AlkD catalyses the hydrolysis of the N-glycosidic bond to liberate an abasic site and free nucleobase. The enzyme is specific for positively charged N3-methyladenine (**a**) and N7-methylguanine (**b**). **c**, **d**, Structures of 3-deaza-3-methyladenosine (**c**) and tetrahydrofuran (**d**) used to trap AlkD in complex with alkylated and abasic DNA. **e**, Crystal structure of AlkD bound to 3d3mA-DNA. Each of the six HEAT repeats is coloured red-to-violet. The DNA is coloured silver with the 3d3mA nucleotide coloured magenta.

<sup>1</sup>Department of Biological Sciences and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37232, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA. <sup>3</sup>Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA.

is comprised entirely of HEAT repeats—tandem  $\alpha$ -helical pairs that generate extended, non-enzymatic scaffolds that typically mediate protein, but not nucleic acid, interactions within their concave surfaces<sup>20–22</sup>. AlkD's concave surface contains highly conserved residues important for 7mG excision and DNA binding activities and protection against bacterial sensitivity to alkylating agents<sup>17–19</sup>.

To investigate the mechanisms by which this novel enzyme binds DNA and catalyses base excision, we determined crystal structures of *B. cereus* AlkD in complex with DNAs resembling the substrate and product of 3mA excision (Fig. 1a). Trapping an alkylpurine DNA glycosylase onto a 3mA-containing substrate has presented a formidable challenge owing to the inherent instability of the *N*-glycosidic bond. To overcome this obstacle, we crystallized AlkD in complex with DNA containing 3-deaza-3-methyladenine (3d3mA), a structural 3mA mimetic in which the N3 nitrogen is replaced with carbon (Fig. 1c). The 3d3mA base is refractory to spontaneous depurination or excision by AlkD or human alkyladenine DNA glycosylase (AAG)<sup>7</sup>, presumably because the 3d3mA purine ring lacks the formal positive charge associated with 3mA. Importantly, the N3→C3 substitution does not affect duplex stability (Supplementary Information)<sup>23</sup>. We also crystallized AlkD in complex with DNA containing a tetrahydrofuran (THF) moiety (Fig. 1d), which resembles the abasic site product. The AlkD–3d3mA–DNA and AlkD–THF–DNA structures were determined by molecular replacement and refined to 1.6 Å ( $R/R_{\text{free}} = 15.9\%/18.3\%$ ) and 1.75 Å ( $R/R_{\text{free}} = 18.5\%/22.5\%$ ), respectively (Supplementary Table 1 and Supplementary Fig. 2).

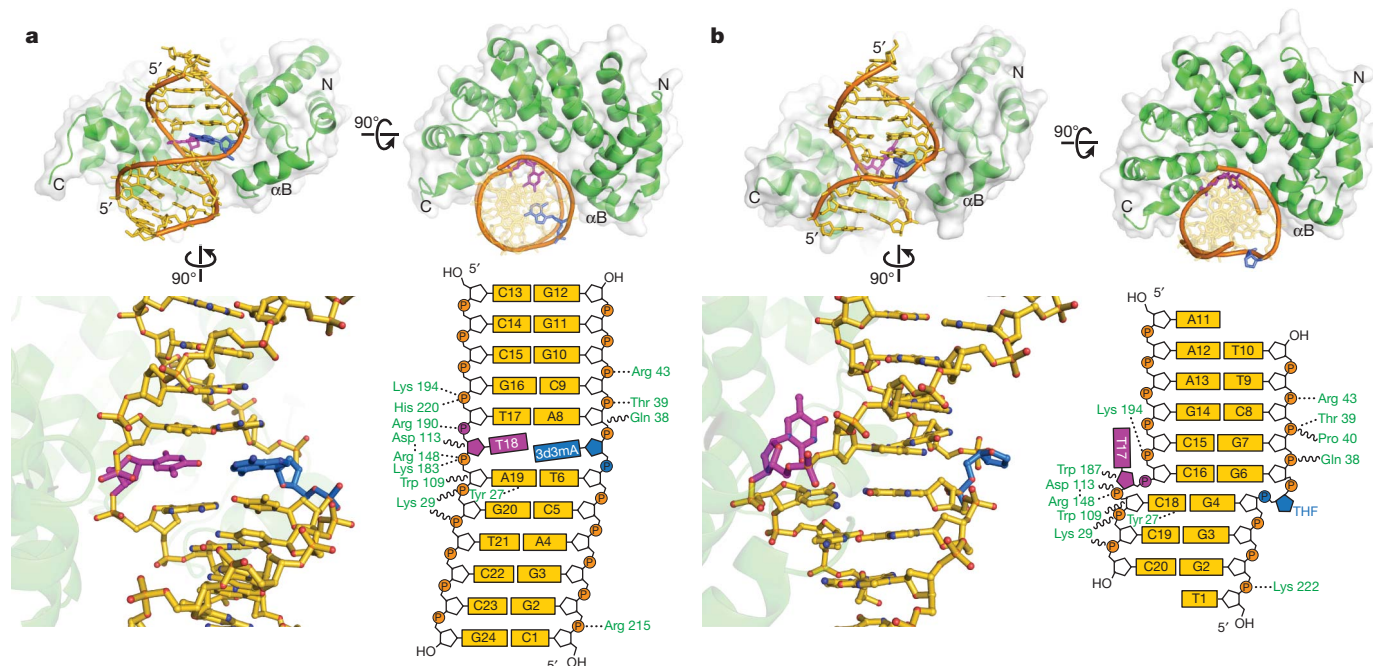
Both 3d3mA and THF complexes show the same general mode of nucleic acid binding despite their unique DNA sequences and crystal packing arrangements (Fig. 2). The DNA is positioned along AlkD's concave surface, which is lined with positively charged residues from the carboxy-terminal  $\alpha$ -helix of each HEAT repeat (Fig. 1e and Supplementary Fig. 3). The C-shaped protein wraps halfway around the DNA helix with a footprint of  $\sim 10$  bp. The contact surface is dominated by electrostatic interactions between side chains at the protein mid-region and the phosphoribose backbone of the DNA strand opposite the lesion. In contrast, contacts to the lesioned strand

are limited to base pairs further removed from the lesion and the protein termini (Fig. 2). The DNA axes are bent  $30^\circ$  away from AlkD's amino terminus as a result of helix  $\alpha$ B (the only non-HEAT repeat in AlkD) projecting into the minor groove (Fig. 2). A 2 Å shift in helix  $\alpha$ B is the only noticeable movement in the protein upon DNA binding (Supplementary Fig. 4).

### A novel lesion capture mechanism

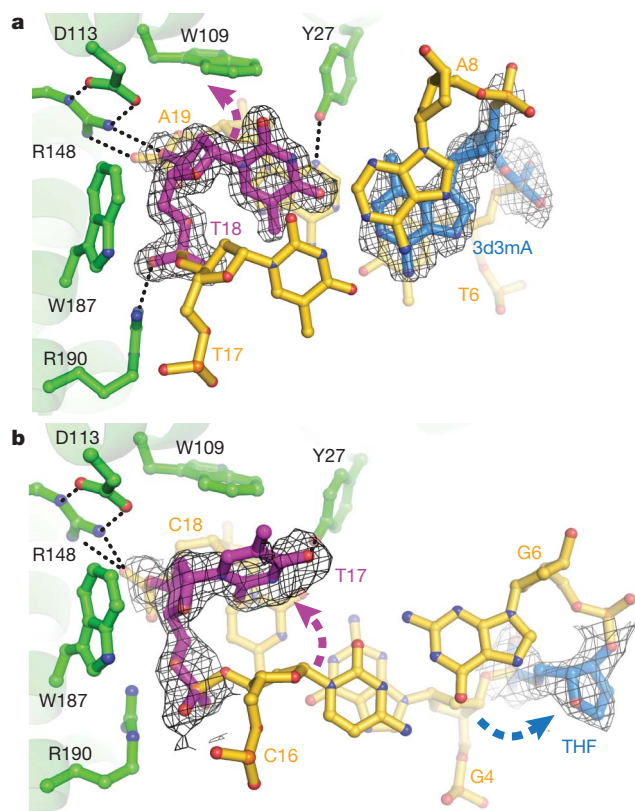
The most striking feature of the AlkD–DNA complexes is that both 3d3mA and THF reside on the face of the DNA duplex not in contact with the protein, whereas the base opposite the lesion is nestled into a cleft on the protein's concave surface (Figs 2 and 3). The 3d3mA•T unpredictably forms a highly sheared base pair in which 3d3mA remains stacked between T6 and A8, whereas the opposite thymine (T18) is displaced into the minor groove with no hydrogen bonds to 3d3mA (Fig. 3a). There are no protein contacts to the T18 base. Rather, it is held in this position by distortion of the T18/A19 backbone as a result of a hydrogen bond network among Asp 113–Arg 148 and Arg 190. The protein–DNA interface is further strengthened by van der Waals interactions between tryptophans 109 and 187 and the phosphoribose backbone flanking the damaged base pair.

In the product complex, the abasic site is rotated  $\sim 90^\circ$  around the phosphoribose backbone into the major groove, and is fully solvent exposed (Figs 2b and 3b). Interestingly, the opposite thymine is slipped completely out of the base stack and into the minor groove of the DNA, and is rotated ( $\chi = 58^\circ$ ) so that the plane of the pyrimidine ring is virtually parallel with the helical axis. Unlike other DNA glycosylases, there is no intercalating side chain plugging the gap left by the flipped base. As a consequence, the duplex has collapsed to maintain base stacking interactions. Guanine G4, immediately 5' to the THF, is now stacked with cytosine C16 on the opposite strand (Fig. 3b). Importantly, the DNA backbone is highly distorted as a result of the large slide (4.4 Å) and twist ( $58^\circ$ ) between G4•C18 and G6•C16 base pairs (Figs 2b and 3b and Supplementary Fig. 12). A hydrogen bond between Tyr 27 at the C-terminal end of helix  $\alpha$ B and the base 3' to the tipped thymine is the only specific AlkD–nucleobase contact (Fig. 3).



**Figure 2 | Crystal structures of AlkD in complex with 3d3mA–DNA (a) and THF–DNA (b).** The top of each panel shows orthogonal views of the AlkD protein (green) wrapping around the DNA duplex (gold). The modified 3d3mA and THF nucleotides are coloured blue, and the opposing thymine is

magenta. At the bottom, a side view of the atomic model and corresponding schematic illustrates the interactions between the modified base pairs and the protein. Dashed lines represent hydrogen bonds and wavy lines represent van der Waals interactions.

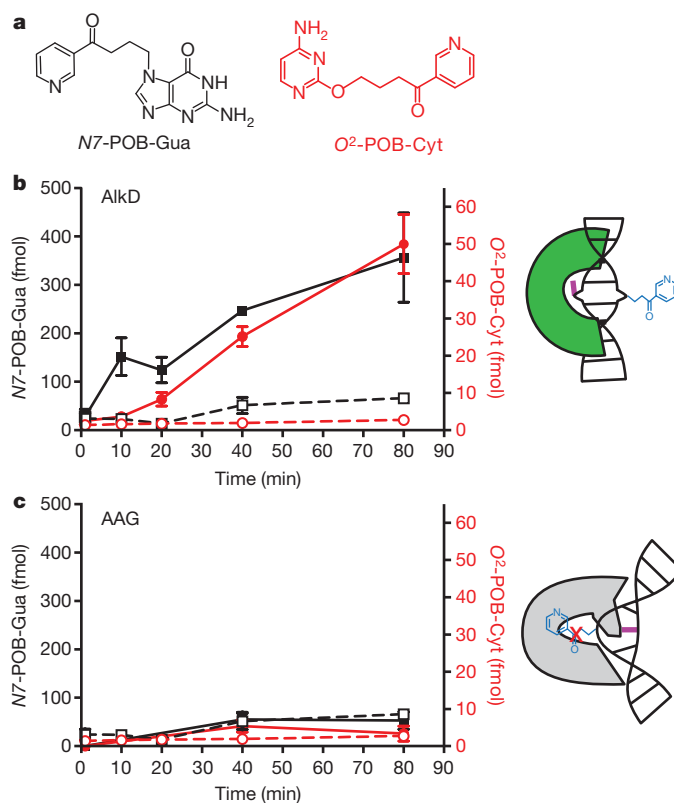


**Figure 3 | Recognition of DNA damage by AlkD.** **a**, 3d3mA-DNA (substrate) complex. **b**, THF-DNA (product) complex. Composite omit electron density (contoured to  $1\sigma$ ) for the modified base pairs is superimposed against the crystallographic models. Dashed arrows denote displacement of THF and opposing thymine from their positions in B-DNA. Hydrogen bonds are shown as dashed lines. Views are down the DNA helix axis.

Thus, AlkD stabilizes the distortions in both substrate and product DNA—a sheared 3d3mA•T base pair and a single base THF•T bubble—through interactions with the phosphoribose backbone of the non-lesioned strand.

The solvent-exposed capture of DNA damage in the AlkD-DNA structures is both unexpected and unprecedented for a DNA glycosylase, and raises the possibility that either AlkD uses a different mechanism to catalyse base excision or that the crystal structures represent nonspecific, catalytically incompetent protein-DNA complexes. Indeed, the aromatic region at the centre of the concave cleft loosely resembles nucleobase binding pockets of other alkylpurine DNA glycosylases<sup>18,19</sup>. However, several important differences argue against a traditional lesion binding pocket in AlkD. First, AlkD lacks the plug residue universally used by DNA glycosylases to prevent the flipped substrate base from re-entering the DNA base stack. Second, an extrahelical nucleobase would be sterically prohibited from full  $180^\circ$  rotation into this shallow cleft (Fig. 3). Third, high concentrations of free nucleobases do not inhibit base excision activity by AlkD as observed in other alkylpurine glycosylases (Supplementary Fig. 5)<sup>15</sup>. Fourth, the electrostatic interaction between Asp 113 and Arg 148 reduces the likelihood that Asp 113 acts as a general base in catalysis. Fifth, mutation of a putative base binding cleft directly adjacent to the catalytic Asp 113 and Arg 148 did not affect 7mG excision activity (Supplementary Fig. 6). Finally, whereas alkylpurine DNA glycosylases normally exhibit enhanced excision activity for mispaired alkylbases, presumably because of their greater propensity to base flip<sup>15,24,25</sup>, AlkD does not discriminate against the base opposite the lesion (Supplementary Table 2 and Supplementary Fig. 7).

To determine the orientation of DNA relative to the central cleft during catalysis, we measured the rate of 7mG excision opposite a



**Figure 4 | Excision of  $N7$ - and  $O^2$ -pyridyloxobutyl (POB) base adducts by AlkD.** **a**, Chemical structures of  $N7$ -POB-guanine and  $O^2$ -POB-cytosine. **b**, **c**, Time courses for the release of  $N7$ -POB-Gua (black squares) and  $O^2$ -POB-Cyt (red circles) in the presence (filled symbols, solid lines) and absence (open symbols, dashed lines) of *B. cereus* AlkD (**b**) or human AAG (**c**). Error bars represent the standard deviation from three independent measurements.

bulky nucleotide. A pyrene nucleotide wedge across from uracil has been shown to enhance base excision by uracil DNA glycosylase (UDG) and rescue the loss of activity of UDG mutants that lack the Leu 191 plug side chain<sup>26</sup>. In contrast, placing pyrene across from 7mG reduced AlkD's activity tenfold relative to a 7mG•C pair (Supplementary Fig. 7). Superposition of the pyrene onto the opposite thymine in the AlkD-DNA crystal structures showed that this bulky group would be hindered from rotating into this tipped position. Thus, the consistency between the crystal structures and partial inhibition of 7mG activity by an opposing pyrene argue strongly that the crystal structures represent a catalytically competent orientation of DNA.

In a converse experiment, we tested the ability of AlkD to excise bulky pyridyloxobutyl (POB) base adducts (Fig. 4a), which arise in DNA upon exposure to cigarette-smoke-derived nitrosamine carcinogens<sup>27</sup>. The expectation was that AlkD should excise POB bases from DNA, whereas the tightly constrained nucleobase binding pocket of human AAG would discriminate against bulky alkyl adducts<sup>28</sup>. Indeed, AlkD liberated positively charged  $N7$ -POB-Gua and  $O^2$ -POB-Cyt adducts from DNA, whereas neither of these modified bases was detected after treatment with AAG or in a mock reaction containing no enzyme (Fig. 4b, c). Neutral adducts  $O^6$ -POB-Gua and  $O^2$ -POB-Thy present in the DNA were not detected in the supernatant upon reaction with AlkD, consistent with the specificity of AlkD for positively charged lesions. This result indicates that AlkD need not flip the substrate base into an active site cavity to excise  $N3$ - or  $N7$ -alkylpurines from DNA.

### AlkD traps and restructures destabilized base pairs

Recent work suggests that DNA glycosylases and oxidative demethylases detect damage by using side chains to probe for free energy



differences between normal and modified base pairs<sup>29–32</sup>. The lack of lesion-specific and DNA intercalating interactions in the AlkD–DNA complexes implies that AlkD detects damage solely on the basis of DNA duplex destabilization resulting from altered stacking or pairing of non-canonical base pairs. In support of this, we crystallized the protein in complex with DNA containing a G•T mismatch (Fig. 5a and Supplementary Table 1), for which AlkD has no activity, but were unable to trap the protein onto the same oligonucleotide containing a G•C or A•T base pair at this same position. The resulting 1.5 Å AlkD–G•T–DNA structure is virtually identical to the 3d3mA•T complex (Supplementary Table 1 and Supplementary Fig. 4). The similarity in these structures, together with thermodynamic differences between modified and unmodified nucleobases, indicates that AlkD detects these energetic differences as opposed to specifically recognizing the N3- or N7-methyl groups (see Supplementary Information)<sup>23,33–36</sup>.

Comparison of the G•T mismatch bound by AlkD and in the context of DNA alone provides a basis for DNA damage recognition by AlkD (Fig. 5). In DNA, G•T wobble mismatches form two Watson–Crick hydrogen bonds and are well stacked within the duplex<sup>37</sup> (Supplementary Fig. 10). AlkD restructures the G•T wobble so that the two bases protrude into opposite DNA grooves, disrupting base stacking and leaving only a single hydrogen bond between guanine N2 and thymine O4 (Fig. 5a). Superposition of a canonical G•T wobble onto the AlkD structure revealed that the protein stabilizes this conformation by inducing a specific distortion to the DNA backbone to alleviate steric clashes (Fig. 5b) and to create optimal hydrogen bonding and van der Waals interactions at the DNA capture site (Supplementary Fig. 10). Thus, the enzyme detects non-Watson–Crick base pairs by resculpting the DNA backbone to create an optimized protein–DNA binding surface. In both 3d3mA•T and G•T complexes, specific protein–DNA contacts are mediated by Arg 148–Asp 113 and Arg 190. Substitution of any of these highly conserved residues reduces single-turnover rates of 7mG excision by an order of magnitude (Fig. 5c), highlighting the importance of these interactions to catalysis.

### Base excision by solvent exposure

The specific structure of the DNA trapped in the AlkD complexes provides a rationale for the enzyme's specificity towards bases with a high propensity for depurination. We propose that the lesion capture mechanism facilitates base hydrolysis by increasing the lifetime that the N-glycosidic bond is exposed to solvent, consistent with spontaneous depurination rates of 7mG in different DNA secondary structural contexts (see Supplementary Information). However, the 100-fold rate enhancement of 7mG hydrolysis from duplex DNA by AlkD cannot be explained on the basis of solvent exposure alone. Close inspection of the highly distorted DNA backbone in the flipped abasic structure revealed

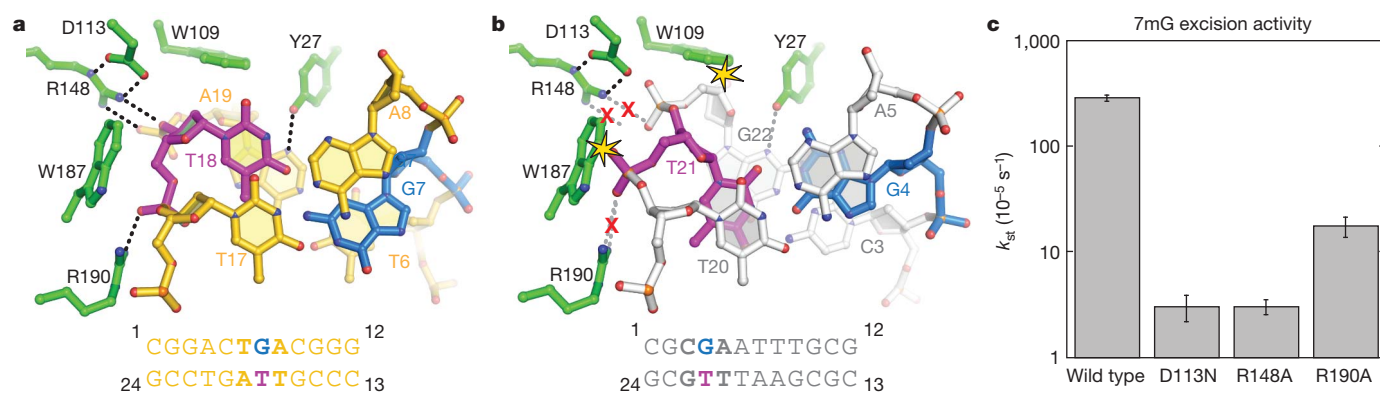
that the deoxyribose ring is positioned directly above a neighbouring phosphate and that several water molecules bridge this phosphate and the extrahelical deoxyribose C1' carbon (Supplementary Fig. 12a), raising the possibility that the phosphate groups participate in catalysis. DNA-mediated water positioning to facilitate hydrolysis is a plausible catalytic mechanism given the lack of a requirement for a general base in these probably highly dissociative reactions. Alternatively, electrostatic stabilization of an oxocarbenium intermediate by nearby phosphates, which has been reported for uracil DNA glycosylase<sup>38,39</sup>, offers a second possible mechanism for DNA-stimulated catalysis.

### Discussion

AlkD represents a novel glycosylase found in bacteria, archaea, plants and eukaryotes (Supplementary Fig. 1)<sup>17</sup>. To our knowledge, most if not all of these organisms contain at least one other alkylpurine DNA glycosylase, raising the question as to why an alternative mechanism has evolved to eliminate genomic alkylation damage. The redundancy of alkylation repair may provide enhanced protection to organisms faced with an onslaught of methylating agents. Alternatively, AlkD may be a general DNA binding protein that coincidentally accelerates hydrolysis of unstable N-glycosidic bonds, or, as speculated below, may have a supporting role in general lesion detection.

AlkD's activity towards bulky POB-DNA adducts normally associated with nucleotide excision repair<sup>40,41</sup> may be indicative of a more generalized function of AlkD in genome maintenance. AlkD's lesion capture strategy is reminiscent of Rad4/XPC, which recognizes cyclopyrimidine dimers by binding to the opposing nucleotides<sup>42</sup>. Exposure of the lesion away from the protein has the biological advantage of damage accessibility by the rest of the nucleotide excision repair machinery. The AlkD–product complex may provide a platform for recruitment of a protein against the extrahelical abasic site, as seen in human APE1–DNA complexes<sup>43</sup>. It is intriguing to speculate that AlkD may participate in alternative repair pathways by virtue of its ability to expose DNA damage. Indeed, non-enzymatic alkyltransferase-like proteins were recently found to trigger nucleotide excision repair of O<sup>6</sup>-alkylguanines by inducing a specific protein–DNA complex, as illustrated by the crystal structure of ATL bound to DNA containing O<sup>6</sup>-POB-dG<sup>44</sup>.

The AlkD–DNA structures illustrate how HEAT repeats engage nucleic acids and, to our knowledge, provide the first structural example of a HEAT motif with enzymatic activity. Comparison with nuclear import factor importin  $\beta$ , which uses HEAT repeats to bind a highly charged region of importin  $\alpha$ <sup>21</sup> and Ran GTPase<sup>22</sup>, demonstrates that the concave surface of the HEAT domain is a generalized macromolecular binding platform. HEAT repeats have been identified in chromatin-remodelling factors, including condensins, cohesins and some SWI2/SNF2 proteins<sup>45</sup>, as well as DNA-damage-response protein



**Figure 5 | Remodelling of a G•T wobble base pair by AlkD.** **a**, AlkD–G•T–DNA complex viewed down the helical axis. AlkD is in green. **b**, The structure of a G•T wobble base pair in DNA alone (Protein Data Bank ID 113D) is superimposed onto the AlkD–G•T complex. Steric clashes between the protein and DNA are highlighted by yellow stars, and disrupted hydrogen bonds

(dashed lines) are shown by a red X. **c**, Relative single-turnover rates ( $k_{st}$ ) of 7mG excision from a 25mer oligonucleotide duplex by wild-type AlkD and the indicated AlkD mutants. Wild-type, D113N and R148A data are from ref. 19. Error bars represent the standard deviation from three independent measurements.

kinases ATM, ATR and DNA-PK<sup>46</sup>. Recently, HEAT domains were visualized by electron microscopy and crystal structures of the catalytic subunit of DNA-PK<sup>47,48</sup>, raising the possibility that other structurally uncharacterized DNA processing enzymes use HEAT domains to bind DNA in a manner similar to AlkD.

## METHODS SUMMARY

**Preparation of 3-deaza-3-methyladenine.** The 3-methyl-3-deazaadenine phosphoramidite was prepared as previously described<sup>49</sup>. The 3-methyl-3-deazaadenine-modified deoxynucleotide oligomers were synthesized at the University of Pittsburgh DNA core facility, purified by reverse-phase HPLC, desalted on Sephadex G20 and analysed by MALDI-TOF-MS. All other oligonucleotides were synthesized by Integrated DNA Technologies.

**AlkD-DNA crystal structure determination.** Wild-type and mutant *B. cereus* AlkD proteins were purified as described previously<sup>19</sup>. AlkD-DNA complexes were assembled using a 1:1.2 molar ratio of AlkD:DNA. AlkD-THF-DNA crystals were grown at 16 °C by vapour diffusion against a reservoir containing 0.1 M Bis-Tris pH 6.5, 0.1 mM NaCl and 9% PEG 3350, and were flash frozen in a 30% glycerol/reservoir solution. AlkD-3d3mA-DNA and AlkD-G•T-DNA crystals were grown at 21 °C from reservoir solutions containing 85 mM NaAcetate pH 4.6, 170 mM ammonium acetate, 25.5% PEG 4000, and 15% glycerol, and were flash frozen directly from the mother liquor. X-ray data (Supplementary Table 1) were collected at the Advanced Photon Source (21-ID-D, LS-CAT). All structures were determined by molecular replacement using the unliganded AlkD structure (Protein Data Bank ID 3BVS) as a search model.

**Biochemical assays.** Base excision and DNA binding activity assays were performed as previously described<sup>19</sup>. Kinetic data were analysed by standard single-turnover techniques. Activity towards POB-nucleobases was measured by incubation of NNK-treated genomic DNA with AlkD or AAG, followed by DNA precipitation and mass-spectrometric detection of N7-POB-Gua and O<sup>2</sup>-POB-Cyt in the supernatant.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 20 November 2009; accepted 17 August 2010.

Published online 3 October 2010.

- Friedberg, E. C. *et al.* DNA repair: from molecular mechanism to human disease. *DNA Repair (Amst.)* **5**, 986–996 (2006).
- Singer, B. & Grunberger, D. *Molecular Biology of Mutagens and Carcinogens: Intrinsic Properties of Nucleic Acids* (Plenum, 1983).
- Holt, S., Yen, T. Y., Sangaiah, R. & Swenberg, J. A. Detection of 1,N<sup>6</sup>-ethenoadenine in rat urine after chloroethylene oxide exposure. *Carcinogenesis* **19**, 1763–1769 (1998).
- Shuker, D. E., Bailey, E., Parry, A., Lamb, J. & Farmer, P. B. The determination of urinary 3-methyladenine in humans as a potential monitor of exposure to methylating agents. *Carcinogenesis* **8**, 959–962 (1987).
- Shuker, D. E. & Farmer, P. B. Relevance of urinary DNA adducts as markers of carcinogen exposure. *Chem. Res. Toxicol.* **5**, 450–460 (1992).
- Larson, K., Sahm, J., Shenkar, R. & Strauss, B. Methylation-induced blocks to *in vitro* DNA replication. *Mutat. Res.* **150**, 77–84 (1985).
- Plosky, B. S. *et al.* Eukaryotic Y-family polymerases bypass a 3-methyl-2'-deoxyadenosine analog *in vitro* and methyl methanesulfonate-induced DNA damage *in vivo*. *Nucleic Acids Res.* **36**, 2152–2162 (2008).
- Gates, K. S., Noonan, T. & Dutta, S. Biologically relevant chemical reactions of N7-alkylguanine residues in DNA. *Chem. Res. Toxicol.* **17**, 839–856 (2004).
- Stivers, J. T. Site-specific DNA damage recognition by enzyme-induced base flipping. *Prog. Nucleic Acid Res. Mol. Biol.* **77**, 37–65 (2004).
- Stivers, J. T. Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chemistry* **14**, 786–793 (2008).
- Stivers, J. T. & Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **103**, 2729–2760 (2003).
- Parikh, S. S. *et al.* Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl Acad. Sci. USA* **97**, 5083–5088 (2000).
- Mol, C. D., Arvai, A. S., Begley, T. J., Cunningham, R. P. & Tainer, J. A. Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases. *J. Mol. Biol.* **315**, 373–384 (2002).
- Drohat, A. C., Kwon, K., Krosky, D. J. & Stivers, J. T. 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member. *Nature Struct. Biol.* **9**, 659–664 (2002).
- Eichman, B. F., O'Rourke, E. J., Radicella, J. P. & Ellenberger, T. Crystal structures of 3-methyladenine DNA glycosylase MagIII and the recognition of alkylated bases. *EMBO J.* **22**, 4898–4909 (2003).
- Metz, A. H., Hollis, T. & Eichman, B. F. DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG). *EMBO J.* **26**, 2411–2420 (2007).
- Alseth, I. *et al.* A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol. Microbiol.* **59**, 1602–1609 (2006).
- Dalhus, B. *et al.* Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. *Nucleic Acids Res.* **35**, 2451–2459 (2007).
- Rubinson, E. H., Metz, A. H., O'Quin, J. & Eichman, B. F. A new protein architecture for processing alkylation damaged DNA: the crystal structure of DNA glycosylase AlkD. *J. Mol. Biol.* **381**, 13–23 (2008).
- Andrade, M. A. & Bork, P. HEAT repeats in the Huntington's disease protein. *Nature Genet.* **11**, 115–116 (1995).
- Cingolani, G., Petosa, C., Weis, K. & Muller, C. W. Structure of importin-β bound to the IBB domain of importin-α. *Nature* **399**, 221–229 (1999).
- Vetter, I. R., Arndt, A., Kutay, U., Gorlich, D. & Wittinghofer, A. Structural view of the Ran-Importin β interaction at 2.3 Å resolution. *Cell* **97**, 635–646 (1999).
- Ganguly, M., Wang, R.-W., Marky, L. A. & Gold, B. Thermodynamic characterization of DNA with 3-deazaadenine and 3-methyl-3-deazaadenine substitutions. *J. Phys. Chem. B* **114**, 7656–7661 (2010).
- O'Brien, P. J. & Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **279**, 9750–9757 (2004).
- O'Brien, P. J. & Ellenberger, T. The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J. Biol. Chem.* **279**, 26876–26884 (2004).
- Jiang, Y. L., Kwon, K. & Stivers, J. T. Turning on uracil-DNA glycosylase using a pyrene nucleotide switch. *J. Biol. Chem.* **276**, 42347–42354 (2001).
- Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat. Res.* **424**, 127–142 (1999).
- Lau, A. Y., Schärer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258 (1998).
- Banerjee, A., Santos, W. L. & Verdine, G. L. Structure of a DNA glycosylase searching for lesions. *Science* **311**, 1153–1157 (2006).
- Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
- Yang, C. G., Garcia, K. & He, C. Damage detection and base flipping in direct DNA alkylation repair. *ChemBioChem* **10**, 417–423 (2009).
- Yang, C. G. *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961–965 (2008).
- Aboul-ela, F., Koh, D., Tinoco, I. Jr & Martin, F. H. Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A, C, G, T). *Nucleic Acids Res.* **13**, 4811–4824 (1985).
- Ezaz-Nikpay, K. & Verdine, G. L. Aberrantly methylated DNA: site-specific introduction of N-7-methyl-2'-deoxyguanosine into the Dickerson/Drew dodecamer. *J. Am. Chem. Soc.* **114**, 6562–6563 (1992).
- Ezaz-Nikpay, K. & Verdine, G. L. The effects of N7-methylguanine on duplex DNA structure. *Chem. Biol.* **1**, 235–240 (1994).
- Lee, S., Bowman, B. R., Ueno, Y., Wang, S. & Verdine, G. L. Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine. *J. Am. Chem. Soc.* **130**, 11570–11571 (2008).
- Hunter, W. N. *et al.* The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.* **262**, 9962–9970 (1987).
- Dinner, A. R., Blackburn, G. M. & Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **413**, 752–755 (2001).
- Jiang, Y. L., Ichikawa, Y., Song, F. & Stivers, J. T. Powering DNA repair through substrate electrostatic interactions. *Biochemistry* **42**, 1922–1929 (2003).
- Brown, P. J., Bedard, L. L. & Massey, T. E. Repair of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced DNA pyridyloxobutylations by nucleotide excision repair. *Cancer Lett.* **260**, 48–55 (2008).
- Li, L. *et al.* The influence of repair pathways on the cytotoxicity and mutagenicity induced by the pyridyloxobutylations pathway of tobacco-specific nitrosamines. *Chem. Res. Toxicol.* **22**, 1464–1472 (2009).
- Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575 (2007).
- Mol, C. D., Izumi, T., Mitra, S. & Tainer, J. A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature* **403**, 451–456 (2000).
- Tubbs, J. L. *et al.* Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* **459**, 808–813 (2009).
- Neuwald, A. F. & Hirano, T. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Res.* **10**, 1445–1452 (2000).
- Perry, J. & Kleckner, N. The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell* **112**, 151–155 (2003).
- Williams, D. R., Lee, K. J., Shi, J., Chen, D. J. & Stewart, P. L. Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals α helices and insight into DNA binding. *Structure* **16**, 468–477 (2008).
- Sibanda, B. L., Chirgadze, D. Y. & Blundell, T. L. Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature* **463**, 118–121 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Stivers for providing the pyrene phosphoramidite, Z. Warzak and LS-CAT beamline staff at the Advanced Photon Source (APS) for assistance with X-ray data collection, and T. Ellenberger, J. Stivers and P. O'Brien for

comments on the manuscript. Use of the APS was supported by the US Department of Energy Office of Basic Energy Sciences. Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor. This research was supported by a grant from the American Cancer Society (to B.F.E.) and the NIH (RO1 CA29088 to B.G.). E.H.R. was supported in part by the Vanderbilt Training Program in Molecular Toxicology. Additional support for local crystallography facilities was provided by the Vanderbilt Center in Molecular Toxicology and the Vanderbilt-Ingram Cancer Center.

**Author Contributions** E.H.R. purified and crystallized AlkD, determined crystal structures and performed 7mG activity assays; B.G. synthesized 3d3mA

oligonucleotides; A.S.P.G. and T.E.S. performed POB activity assays; B.F.E. designed the project; B.F.E. and E.H.R. analysed data and wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3JX7 (3d3mA•T), 3JXY (G•T), 3JXZ (THF•T) and 3JY1 (THF•C). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to B.F.E. ([brandt.eichman@vanderbilt.edu](mailto:brandt.eichman@vanderbilt.edu)).



## METHODS

**AlkD purification and crystallization.** AlkD proteins were purified as described previously<sup>19</sup>. Briefly, *Bacillus cereus* AlkD was overexpressed as an N-terminal His<sub>6</sub>-SUMO-AlkD fusion protein in *E. coli* HMS174 cells for 3 h at 37 °C. AlkD was isolated using Ni-NTA (Qiagen) affinity chromatography, followed by cleavage of the His<sub>6</sub>-SUMO tag and further purification by heparin affinity and gel filtration chromatography. Protein was concentrated to 12.5 mg ml<sup>-1</sup> and stored in 20 mM Bis-Tris propane, 100 mM NaCl, 2 mM DTT and 0.1 mM EDTA. Site-directed mutagenesis of the wild-type AlkD vector was performed using a Quik-Change Kit (Stratagene). Mutant proteins were overexpressed and purified identically to wild-type AlkD, and their structures verified by circular dichroism spectroscopy.

AlkD–DNA complexes were assembled by incubating 0.45 mM protein and 0.54 mM oligonucleotide for 15 min at 4 °C. Oligonucleotide sequences used were d(TGGG(THF)GGCTT)/d(AAAGCCYCCC), in which Y = T or C, and d(CGGACTXACGGG)/d(CCCGTTTCCG), in which X = 3d3mA<sup>49</sup> or G. AlkD–THF–DNA crystals were grown at 16 °C by mixing 2 µl protein–DNA complex with 2 µl reservoir solution containing 0.1 M Bis-Tris pH 6.5, 0.1 mM NaCl, and 19% PEG 3350 and 2% glycerol. Crystals were soaked in 30% glycerol/reservoir solution for 1 min and flash frozen in a liquid nitrogen stream. Crystals of 3d3mA–DNA and G•T–DNA complexes were grown from reservoir solutions containing 85 mM NaAcetate pH 4.6, 170 mM ammonium acetate, 25.5% PEG 4000, and 15% glycerol at 21 °C, and were flash frozen in liquid nitrogen directly from this solution.

**X-ray data collection, phasing and structure refinement.** X-ray data (Supplementary Table 1) were collected at a wavelength of 0.97850 Å and 110 K at the Advanced Photon Source beamlines 21-ID-D and 21-ID-G (LS-CAT) and processed with HKL2000<sup>50</sup>. Molecular replacement using unliganded AlkD (Protein Data Bank ID 3BVS) as a search model in Phaser<sup>51</sup> gave a clear solution for each structure. After one round of simulated annealing refinement in CNS<sup>52</sup>, the entire DNA molecules could be discerned and were built into 2F<sub>o</sub> – F<sub>c</sub> and F<sub>o</sub> – F<sub>c</sub> electron density using XtalView<sup>53</sup>. Atomic coordinates and B-factors for the protein–DNA models were refined in Phenix<sup>54</sup>. TLS refinement with protein and each DNA chain defined as three separate TLS groups was carried out for each model except the GT-complex. Individual anisotropic B-factors were derived from the refined TLS parameters and held fixed during subsequent rounds of refinement, which significantly decreased the crystallographic residuals and improved the electron density maps. Instead of TLS refinement, individual anisotropic B-factors were explicitly refined for the G•T complex. Adjustments to the model, including addition of solvent molecules, using Coot<sup>55</sup> were guided by manual inspection of 2F<sub>o</sub> – F<sub>c</sub> and F<sub>o</sub> – F<sub>c</sub> electron density maps and were judged successful by a decrease in R<sub>free</sub> during refinement.

Protein and DNA models were validated using PROCHECK<sup>56</sup> and DNA parameters were quantified using CURVES 5.2<sup>57</sup>. All but one out of the total 223–231 protein residues resided in the most favoured (191–198 residues) or allowed (14–15 residues) regions of the Ramachandran plot. As in the unliganded structure<sup>19</sup>, Thr 54 in all four DNA complex structures remained in the disallowed region despite an excellent fit to 2F<sub>o</sub> – F<sub>c</sub> electron density maps.

**Enzyme activity.** Excision of 7mG by AlkD was measured by incubating the enzyme with a 25mer oligonucleotide containing a centrally located 7mG and following the appearance of abasic DNA product after alkaline cleavage. 7mG was enzymatically incorporated into DNA duplexes using the previously described method<sup>58</sup>, in which an oligonucleotide primer (d(GACCACTACACC)) was <sup>32</sup>P-labelled at the 5' end, annealed to a threefold excess of the complementary strand (d(GTTGTTAGGAAACGGTGTAGTGGTC)) and extended using DNA polymerase I Klenow fragment (New England Biolabs) in the presence of 2'-deoxy-7-methylguanosine 5'-triphosphate (Sigma), dCTP, dTTP and dATP. To create 7mG mispairs, 100-fold excess of complementary strand with T, G, A or pyrene in place of C at position 13 was re-annealed to the 7mG containing oligonucleotide. Single-stranded 7mG containing strands were obtained by re-annealing to 100-fold excess of unlabelled lesion strand with G in place of 7mG (d(GACCACTACACCGTTTCTCAACAAC)).

In a 10 µl glycosylase reaction, 100 nM [<sup>32</sup>P]–DNA duplex was incubated with 0–20 µM AlkD in 50 mM HEPES pH 7.5, 100 mM NaCl, 10 mM DTT and 2 mM EDTA. The reaction was quenched at various times by the addition of 0.2 N NaOH and heated at 70 °C for 2 min. Substrate 25mer and product 12mer DNA strands were separated by denaturing 20% polyacrylamide gel electrophoresis in 7 M urea and quantified by autoradiography. Kinetic data were analysed by standard single-turnover techniques<sup>59</sup>, which have been extensively used for DNA glycosylases<sup>60–65</sup>. Enzymatic rate constants (*k*) were obtained from single-exponential fits to the data (*f<sub>p</sub>* = 1 – e<sup>–*kt*</sup>, in which *f<sub>p</sub>* is the fraction of product). For determination of the single-turnover rate constant, *k<sub>st</sub>*, AlkD was at least fivefold in excess over the *K<sub>1/2</sub>* for a particular labelled DNA substrate (for example, 5 µM for 7mG•C). For *K<sub>1/2</sub>*

determinations, the 7mG excision assay was performed over a range of enzyme concentrations and *K<sub>1/2</sub>* obtained by fitting the Michaelis–Menten plot with the equation, *k<sub>obs</sub>* = *V<sub>max</sub>*[AlkD]/(*K<sub>1/2</sub>* + [AlkD]). We note that our *K<sub>1/2</sub>* for maximal activity may differ from the *K<sub>m</sub>* value for multiple turnover because the *K<sub>m</sub>* can be affected by product release. Stoichiometric 7mG excision was performed in the presence of 10 µM unlabelled 25mer DNA duplex (*K<sub>1/2</sub>* for this DNA was determined to be 0.9 ± 0.1 µM). Spontaneous rates of 7mG hydrolysis were determined using the sequence d(GACCACTACACC(7mG)ATTCCTTACAAC) that had been re-annealed to 100-fold excess complementary strand d(GTTGTAAGG AAT(C/T)GGTGTAGTGGTC).

**POB adduct excision.** DNA (catalogue no. D1501), alkaline phosphatase (P8361), esterase (E2884) micrococcal nuclease (N3755) and phosphodiesterase II (P9041) were purchased from Sigma. The tetra-deuterated standards were provided by S. S. Hecht. NNKOAc was synthesized by D. Desai.

For NNKOAc-damaged DNA, 5 ml DNA (2 mg ml<sup>-1</sup>) dissolved in 100 mM sodium phosphate (pH 7.0), 1 mM EDTA and 50 mM NaCl was reacted with 1 mM NNKOAc and esterase (200 units) at 37 °C for 2 h. The reaction was diluted to 10 ml with H<sub>2</sub>O and extracted with 10 ml CHCl<sub>3</sub>/iso-amyl alcohol (24/1) to remove the protein and 10 ml ethyl acetate to remove any unreacted NNKOAc. The DNA was precipitated by the addition of 40 ml ethanol and washed twice with 70% ethanol. Residual amounts of ethanol were removed by rotary evaporation and the DNA was dissolved in H<sub>2</sub>O, aliquoted and stored at –80 °C before use.

For glycosylase reactions, the damaged DNA (1 mg ml<sup>-1</sup>) was incubated with 1 µM glycosylase in 400 µl buffer (50 mM HEPES (pH 7.5), 1 mM EDTA, 100 mM KCl, 1 mM DTT) at 37 °C. Aliquots (100 µl) were quenched at various times by the addition of 5 µl 3 M sodium acetate (pH 5.2) and 200 µl ice-cold ethanol. The mixture was centrifuged for 10 min and the supernatant decanted and saved for analysis.

For HPLC-MS/MS, deuterated standards (100 fmol each of O<sup>2</sup>-POB-C-d<sub>4</sub> and N7-POB-G-d<sub>4</sub>) were added to the ethanol supernatant and the solvent evaporated. The sample was dissolved in 50 µl methanol for MS analysis. The samples were analysed with a MDS/Sciex 4000 QTrap instrument with electrospray ionization (ESI) coupled to an Agilent 1100 HPLC system. Samples (20 µl) were loaded onto a column (Luna C18(2) 150 × 2 mm, 3 µm) which was eluted with 10 mM ammonium formate at 0.1 ml min<sup>-1</sup>. The POB–DNA adducts, along with their deuterated standards, were monitored by selected reaction monitoring. The ion transitions were as follows N7-POB-Gua, *m/z* 299.1 [M + 1]<sup>+</sup> to *m/z* 148.1 [POB]<sup>+</sup>; [pyridine-D<sub>4</sub>]N7-POB-Gua, *m/z* 303.1 [M + 1]<sup>+</sup> to *m/z* 152.1 ([pyridine-D<sub>4</sub>]POB)<sup>+</sup> and [Gua + H]<sup>+</sup>; O<sup>2</sup>-POB-Cyt, *m/z* 259.1 [M + 1]<sup>+</sup> to *m/z* 148.1 [POB]<sup>+</sup>; [pyridine-D<sub>4</sub>]O<sup>2</sup>-POB-Cyt, *m/z* 263.1 [M + 1]<sup>+</sup> to *m/z* 152.1 ([pyridine-D<sub>4</sub>]POB)<sup>+</sup>. Prior to HPLC-ESI-MS/MS analysis of the samples the MS parameters were optimized for each deuterated POB–DNA adduct standard. For analysis, the MS parameters were set as follows: curtain gas, 40 p.s.i.; ion spray voltage, 4 kV; source temperature, 650 °C; nebulizer gas (GS1), 70 p.s.i.; heater gas (GS2), 70 p.s.i.; and collision gas, 12 p.s.i. The fragmentation potentials were optimized for each ion. For 299.1 and 303.1: declustering potential (DP), 65 V; entrance potential (EP), 10 V; collision energy (CE), 20 V; collision cell exit potential (CXP), 12 V. For 259.1 and 263.1: DP, 40 V; EP, 8 V; CE, 15 V; and CXP, 6 V. The amount of each POB–DNA adduct was determined by comparing the MS peak area ratio of each adduct to its deuterated standard with a calibration curve. Calibration standards were prepared by spiking different amounts of each adduct with a constant amount of the corresponding internal standard in H<sub>2</sub>O and then analysed by LC-MS/MS without undergoing the sample preparation procedure described above. The calibration curves were constructed by plotting concentration ratio versus MS peak area ratios of each adduct to its deuterated standard.

**DNA binding.** DNA binding was monitored by a change in fluorescence anisotropy as increasing concentrations of protein were added to an oligonucleotide duplex that contained a THF abasic modification in the middle of one strand (d(TGACTACTACAT(THF)GTTGCCTACCAT)) and a 6-carboxyfluorescein (FAM) on the 3' end of the complementary strand (d(ATGGTAGGCAACTA TGTAGTAGTCA)-FAM). For stoichiometric binding measurements, increasing concentrations of protein (0–200 µM) were added to a solution containing 50 nM FAM–DNA and 20 µM unlabelled 25mer DNA (*K<sub>d</sub>* = 3.1 ± 0.3 µM) in 20 mM Bis-Tris propane pH 6.5, 100 mM NaCl, 2 mM DTT and 0.1 mM EDTA. Polarized fluorescence intensities using excitation and emission wavelengths of 485 and 538 were measured at ambient temperature using a SpectraMax M5 microplate reader (Molecular Devices). Dissociation constants were derived by fitting a two-state binding model to data from three independent experiments.

49. Irani, R. J. & SantaLucia, J. Jr. The synthesis of anti-fixed 3-methyl-3-deaza-2'-deoxyadenosine and other 3H-imidazo[4,5-c]pyridine analogs. *Nucleosides Nucleotides Nucleic Acids* **21**, 737–751 (2002).

50. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
51. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
52. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
53. McRee, D. E. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
54. Adams, P. D. *et al.* in *Evolving Methods for Macromolecular Crystallography* (eds Read, R. J. & Sussman, J. L.) 101–109 (Springer, 2007).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
56. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
57. Lavery, R. & Sklenar, H. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63–91 (1988).
58. Asaeda, A. *et al.* Substrate specificity of human methylpurine DNA N-glycosylase. *Biochemistry* **39**, 1959–1965 (2000).
59. Jones, B. N., Quang-Dang, D. U., Oku, Y. & Gross, J. D. A kinetic assay to monitor RNA decapping under single-turnover conditions. *Methods Enzymol.* **448**, 23–40 (2008).
60. Baldwin, M. R. & O'Brien, P. J. Human AP endonuclease 1 stimulates multiple-turnover base excision by alkyladenine DNA glycosylase. *Biochemistry* **48**, 6022–6033 (2009).
61. Lyons, D. M. & O'Brien, P. J. Efficient recognition of an unpaired lesion by a DNA repair glycosylase. *J. Am. Chem. Soc.* **131**, 17742–17743 (2009).
62. Maher, R. L. & Bloom, L. B. Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *J. Biol. Chem.* **282**, 30577–30585 (2007).
63. Maher, R. L., Vallur, A. C., Feller, J. A. & Bloom, L. B. Slow base excision by human alkyladenine DNA glycosylase limits the rate of formation of AP sites and AP endonuclease 1 does not stimulate base excision. *DNA Repair (Amst.)* **6**, 71–81 (2007).
64. Maiti, A., Morgan, M. T. & Drohat, A. C. Role of two strictly conserved residues in nucleotide flipping and N-glycosylic bond cleavage by human thymine DNA glycosylase. *J. Biol. Chem.* **284**, 36680–36688 (2009).
65. Bennett, M. T. *et al.* Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **128**, 12510–12519 (2006).

# Entanglement of spin waves among four quantum memories

K. S. Choi<sup>1</sup>, A. Goban<sup>1</sup>, S. B. Papp<sup>1†</sup>, S. J. van Enk<sup>2</sup> & H. J. Kimble<sup>1</sup>

Quantum networks are composed of quantum nodes that interact coherently through quantum channels, and open a broad frontier of scientific opportunities<sup>1</sup>. For example, a quantum network can serve as a ‘web’ for connecting quantum processors for computation<sup>2,3</sup> and communication<sup>4</sup>, or as a ‘simulator’ allowing investigations of quantum critical phenomena arising from interactions among the nodes mediated by the channels<sup>5,6</sup>. The physical realization of quantum networks generically requires dynamical systems capable of generating and storing entangled states among multiple quantum memories, and efficiently transferring stored entanglement into quantum channels for distribution across the network. Although such capabilities have been demonstrated for diverse bipartite systems<sup>7–12</sup>, entangled states have not been achieved for interconnects capable of ‘mapping’ multipartite entanglement stored in quantum memories to quantum channels. Here we demonstrate measurement-induced entanglement stored in four atomic memories; user-controlled, coherent transfer of the atomic entanglement to four photonic channels; and characterization of the full quadripartite entanglement using quantum uncertainty relations<sup>13–16</sup>. Our work therefore constitutes an advance in the distribution of multipartite entanglement across quantum networks. We also show that our entanglement verification method is suitable for studying the entanglement order of condensed-matter systems in thermal equilibrium<sup>17,18</sup>.

Diverse applications in quantum information science require coherent control of the generation, storage and transfer of entanglement among spatially separated physical systems<sup>1–6</sup>. Despite its inherently multipartite nature, entanglement has been studied primarily for bipartite systems<sup>3</sup>, where remarkable progress has been made in harnessing physical processes to generate ‘push-button’ and ‘heralded’ entanglement<sup>7–10,19,20</sup>, as well as to map entangled states to and from atoms, photons and phonons<sup>11,12</sup>.

For multipartite systems, the ‘size’ of a physical state, described by the system’s density matrix,  $\hat{\rho}_N$ , grows exponentially with the number of subsystems,  $N$ , and makes the entangled states exceedingly difficult to represent with classical information. Importantly, this complexity of  $\hat{\rho}_N$  increases the potential utility of multipartite entanglement in quantum information science, including quantum algorithms<sup>2,3</sup> and simulation<sup>5</sup>. Redundant encoding of quantum information into multipartite entangled states allows quantum error correction and fault-tolerant computation<sup>2,3</sup>. Intricate long-range correlation of many-body systems is intimately intertwined with the behaviour of multipartite entanglement<sup>17,18</sup>. In addition, mobilizing multipartite entanglement across quantum networks could lead to novel quantum phase transitions for the network<sup>6</sup>.

Counterposed to these opportunities, the complex structure of multipartite entanglement presents serious challenges both for its formal characterization and physical realization<sup>3,18,21,22</sup>. Indeed, there are relatively few examples of laboratory systems that have successfully generated multipartite entanglement<sup>15,23–27</sup>. Most works have considered the entanglement in spin systems, notably trapped ions<sup>23,24</sup>,

which are applicable to the matter nodes of quantum networks. But the methodologies for verifying multipartite entanglement are problematic for infinite-dimensional bosonic systems of the quantum channels (for example multipartite quadrature<sup>25,26</sup> and number-state<sup>15</sup> entanglement for optical modes). A-posteriori multipartite entanglement has been inferred from a small subset of preferred photon detection events from parametric down-conversion<sup>27</sup>.

In addition to the characterization of multipartite entanglement, an important capability of quantum networks is provided by quantum interfaces capable of generating, storing and dynamically allocating the entanglement of matter nodes into photonic channels (see ref. 28 and references therein). Here we introduce such a quantum interface for quadripartite entangled states based upon coherent, collective emission from matter to light, as illustrated in Fig. 1a. We present a systematic study of the generation and storage of quadripartite entangled states of spin waves in a set of four nodes of atomic memories, as well as of the coherent transfer of the entangled components of the material state into individual photonic channels. We observe transitions of  $M$ -partite to  $(M - 1)$ -partite entangled states via controlled spin-wave statistics of the atomic memories, as well as the dynamic evolution of multipartite entanglement in a dissipative environment, from fully quadripartite entangled states to unentangled states.

Our experiment proceeds in four steps (Methods). First, in step (i), an entangled state,  $\hat{\rho}_W^{(A)}$ , of four atomic ensembles is generated by quantum interference in a quantum measurement<sup>4,7</sup> (Fig. 1b). Given a photoelectric detection event at detector  $D_h$ , the conditional atomic state is ideally a quadripartite entangled state,  $\hat{\rho}_W^{(A)} = |W\rangle_A \langle W|$ , with

$$|W\rangle_A = \frac{1}{2} \left[ (|\bar{s}_a \bar{g}_b \bar{g}_c \bar{g}_d\rangle + e^{i\phi_1} |\bar{g}_a \bar{s}_b \bar{g}_c \bar{g}_d\rangle) + e^{i\phi_2} (|\bar{g}_a \bar{g}_b \bar{s}_c \bar{g}_d\rangle + e^{i\phi_3} |\bar{g}_a \bar{g}_b \bar{g}_c \bar{s}_d\rangle) \right] \quad (1)$$

whose single quantum spin wave,  $|\bar{s}_\varepsilon\rangle$ , is coherently shared among four ensembles,  $\varepsilon \in \{a, b, c, d\}$ . These entangled states are known as W states, and comprise atomic ground states,  $|\bar{g}_\varepsilon\rangle = |g \cdots g\rangle_\varepsilon$ , and single collective excitations,  $|\bar{s}_\varepsilon\rangle = (1/\sqrt{N_{A,\varepsilon}}) \sum_{i=1}^{N_{A,\varepsilon}} |g \cdots s_i \cdots g\rangle_\varepsilon$ , where  $N_{A,\varepsilon}$  is the number of atoms in ensemble  $\varepsilon$ .

After the heralding event, step (ii) consists of storage of  $\hat{\rho}_W^{(A)}$  in the ensembles for a user-controlled time,  $\tau$ . At the end of this interval, step (iii) is initiated with read beams to coherently transfer the entangled atomic components of  $\hat{\rho}_W^{(A)}$  into a quadripartite entangled state of light,  $\hat{\rho}_W^{(\gamma)} = |W\rangle_\gamma \langle W|$ , by means of cooperative emissions<sup>4</sup> (Fig. 1c), where

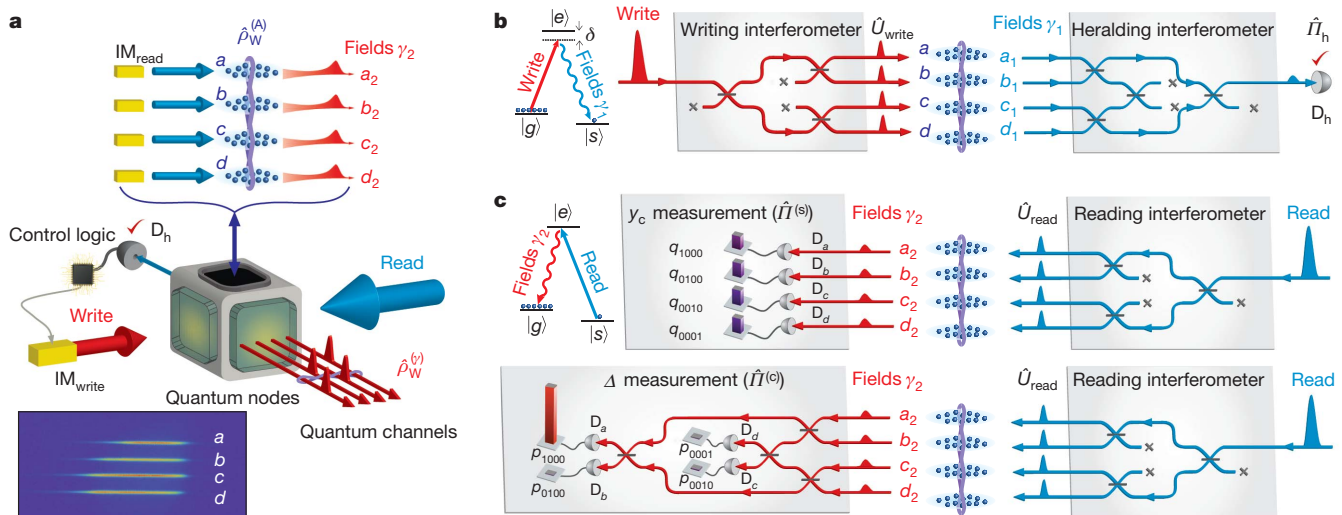
$$|W\rangle_\gamma = \frac{1}{2} \left[ (|1000\rangle + e^{i\phi'_1} |0100\rangle) + e^{i\phi'_2} (|0010\rangle + e^{i\phi'_3} |0001\rangle) \right] \quad (2)$$

This photonic state is a mode-entangled W state<sup>15,16</sup>, which shares a single delocalized photon among four spatially separated optical modes,  $\gamma_2 \in \{a_2, b_2, c_2, d_2\}$ .

Finally, in step (iv) we characterize the heralded entanglement from complementary measurements of photon statistics and

<sup>1</sup>Norman Bridge Laboratory of Physics 12-33, California Institute of Technology, Pasadena, California 91125, USA. <sup>2</sup>Department of Physics, University of Oregon, Eugene, Oregon 97403, USA. <sup>†</sup>Present address: National Institute of Standards and Technology, Boulder, Colorado 80305, USA.





**Figure 1 | Overview of the experiment.** **a**, Quantum interfaces for multipartite quantum networks. Inset, a fluorescence image of the laser-cooled atomic ensembles  $a$ ,  $b$ ,  $c$  and  $d$  that become entangled (Methods).  $\text{IM}_{\text{read}}$  and  $\text{IM}_{\text{write}}$  are the respective intensity modulators of the read and write lasers. **b**, Entanglement generation. A weak write laser (red-detuned by  $\delta = 10$  MHz from the  $|g\rangle \rightarrow |e\rangle$  transition) is split into four components to excite the atomic ensembles by means of parametric interactions,  $\hat{U}_{\text{write}}$ , leading to Raman scattered fields,  $\gamma_1 \in \{a_1, b_1, c_1, d_1\}$ , emitted by the ensembles. The entangled state,  $\hat{\rho}_W^{(A)}$ , for four atomic ensembles  $\varepsilon \in \{a, b, c, d\}$  (equation (1)) is heralded by a projective measurement,  $\hat{\Pi}_h$ , at detector  $D_h$ , derived from quantum interference of four fields  $\gamma_1$  in the heralding interferometer. **c**, Quantum state

exchange and entanglement verification. Read lasers are applied to the ensembles to transform the atomic entangled state  $\hat{\rho}_W^{(A)}$  coherently into quadripartite entangled beams of light,  $\hat{\rho}_W^{(v)}$  (equation (2)) by means of quantum state transfers,  $\hat{U}_{\text{read}}$ , with each beam propagating through quantum channels  $\gamma_2 \in \{a_2, b_2, c_2, d_2\}$ . Subpanel for  $y_c$  measurement: the quantum statistics,  $q_{ijkl}$ , for the individual modes of  $\hat{\rho}_W^{(v)}$  with  $i, j, k, l \in \{0, 1\}$  photons are measured with projectors  $\hat{\Pi}_i^{(s)}$  at detectors  $D_a, D_b, D_c, D_d$ . Subpanel for  $\Delta$  measurement: mutual coherences for  $\hat{\rho}_W^{(A)}$  are accessed with projectors  $\hat{\Pi}_i^{(c)}$  from detection statistics  $p_{ijkl}$  at  $D_a, D_b, D_c, D_d$ . Further details are given in Supplementary Information.

coherence<sup>15,16</sup> (Fig. 1c). In particular, we consider a reduced density matrix,  $\hat{\rho}_r = p_0 \hat{\rho}_0 + p_1 \hat{\rho}_1 + p_{\geq 2} \hat{\rho}_{\geq 2}$ , containing up to one photon per mode, which leads to a lower bound for the entanglement of the actual physical states,  $\hat{\rho}_W^{(A)}$  and  $\hat{\rho}_W^{(v)}$ . Here  $p_0, p_1$  and  $p_{\geq 2}$  are the probabilities of the zero- and one-photon subspaces ( $\hat{\rho}_0$  and  $\hat{\rho}_1$ ) and the higher-order subspaces ( $\hat{\rho}_{\geq 2}$ ), which can be populated for any realistic system. As illustrated in the upper panel of Fig. 1c, we characterize the statistical contamination of  $\hat{\rho}_W^{(v)}$  due to  $\hat{\rho}_0$  and  $\hat{\rho}_{\geq 2}$  with a normalized measure<sup>16</sup>—namely  $y_c \equiv (8/3)p_{\geq 2}p_0/p_1^2$ , which ranges from  $y_c = 0$ , for a single excitation, to  $y_c = 1$ , for balanced coherent states—by detecting the photon statistics,  $q_{ijkl}$ , of  $\gamma_2$  at the output faces of the ensembles.

We also quantify the mutual coherences of  $\hat{\rho}_W^{(v)}$  by measuring the photon probabilities  $p_{1000}, p_{0100}, p_{0010}$  and  $p_{0001}$  at the outputs of the verification (v) interferometer. We determine the sum uncertainty,  $\Delta \equiv \sum_{i=1}^N \left\langle \left( \hat{\Pi}_i^{(c)} \right)^2 - \left\langle \hat{\Pi}_i^{(c)} \right\rangle^2 \right\rangle$  for the variables  $\hat{\Pi}_i^{(c)} = |W_i\rangle_v \langle W_i|$ , which project  $\hat{\rho}_r$  onto a set of four orthonormal W states,  $|W_i\rangle_v$ , with phases,  $\beta_i \in \{\beta_1, \beta_2, \beta_3\}_v$ , selected by the actively stabilized paths in the verification interferometer (Supplementary Information). Hence, for the ideal W state (equation (2)) with  $\beta_i = \phi'_i$ , we have  $\Delta = 0$  associated with  $p_{1000} = 1$  and  $p_{0100} = p_{0010} = p_{0001} = 0$ , as observed in the bar plots of the lower panel of Fig. 1c for  $y_c = 0.04 \pm 0.01$ . In contrast, mixed states with no phase coherences would result in balanced probabilities ( $p_{1000} = p_{0100} = p_{0010} = p_{0001} = 1/4$ ) and  $\Delta = 0.75$ .

The pair  $\{\Delta, y_c\}$  thereby defines the parameter space for the multipartite entanglement in our experiment, with the entanglement parameters  $\Delta$  and  $y_c$  serving as a non-local, nonlinear entanglement witness<sup>16</sup>. Our criterion for ‘genuine’  $M$ -partite entanglement takes the most stringent form of non-separability (ref. 22 and references therein) and excludes all weaker forms of entanglement (Methods). Specifically, for a given value of  $y_c$ , we determine the boundary,  $\Delta_b^{(M-1)}$ , for the minimal uncertainty possible for all states containing at most  $(M-1)$ -mode entanglement and their mixtures (Supplementary Information). For our quadripartite states ( $N = 4$ ), we derive  $\Delta_b^{(3)}$ ,  $\Delta_b^{(2)}$  and  $\Delta_b^{(1)}$  for tripartite entangled, bipartite entangled and fully

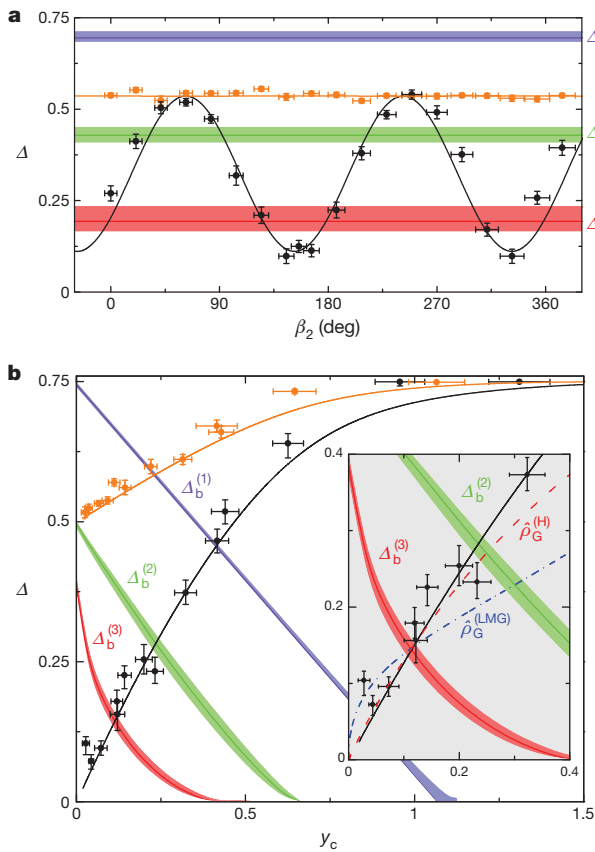
separable states, respectively, as functions of  $y_c$ . Thus, a measurement of quantum statistics ( $y_c$ ) and the associated coherence ( $\Delta$ ) with  $\Delta < \Delta_b^{(3)}, \Delta_b^{(2)}$  and  $\Delta_b^{(1)}$  manifestly confirms the presence of genuine ( $M = 4$ )-partite entanglement<sup>15,16</sup>. Furthermore, we can unambiguously distinguish genuine  $M$ -partite and  $(M-1)$ -partite entangled states for any  $M \leq N$  by observing  $\Delta$  below  $\Delta_b^{(M-1)}$ .

Figure 2 presents our results for quadripartite entanglement for a storage time of  $\tau_0 = 0.2 \mu\text{s}$ . We first investigate off-diagonal coherence for the purportedly entangled atomic and photonic states,  $\hat{\rho}_W^{(A)}$  and  $\hat{\rho}_W^{(v)}$ , in Fig. 2a. As the bipartite phase,  $\beta_2$ , is varied, we observe interferences in  $p_{1000}, p_{0100}, p_{0010}$  and  $p_{0001}$ , and, hence, a variation in  $\Delta$  that results from the coherence between the bipartite entangled components of  $\hat{\rho}_W^{(v)}$  for the modes  $\{a_2, b_2\}$  and  $\{c_2, d_2\}$ . Furthermore, for optimal settings of  $\beta_2$ , the observed values of  $\Delta$  (Fig. 2a, black points) fall below the bounds  $\Delta_b^{(3)}, \Delta_b^{(2)}$  and  $\Delta_b^{(1)}$  (red, green and purple bands, respectively) for  $y_c = 0.06 \pm 0.02$ , and signal the generation of a fully quadripartite entangled state. The observed quadripartite entanglement arises from the intrinsic indistinguishability of probability amplitudes for one collective excitation,  $|\bar{s}_e\rangle$ , among the four ensembles. We also present results from a control experiment with a ‘crossed’ state,  $\hat{\rho}_W^{(A)}$  (Fig. 2a, orange points), that consists of an incoherent mixture of entangled pairs  $\{a, b\}$  and  $\{c, d\}$  (Methods).

Next we characterize  $\hat{\rho}_W^{(v)}$  (and  $\hat{\rho}_W^{(A)}$ ) over the full parameter space,  $\{\Delta, y_c\}$ . In a regime of weak excitation (with excitation probability  $\xi \ll 1$ ) for the ensemble–field pairs  $\{\varepsilon, \gamma_1\}$ , the heralded state  $\hat{\rho}_W^{(A)}$  is approximately

$$\hat{\rho}_W^{(A)}(\tau=0) \approx (1-3\xi)|W\rangle_A \langle W| + 3\xi \hat{\rho}_{\geq 2}^{(A)} + \mathcal{O}(\xi^2) \quad (3)$$

where  $\hat{\rho}_{\geq 2}^{(A)}$  includes uncorrelated spin waves with two or more quanta in the set of four ensembles due to atomic noise. As  $\xi \rightarrow 0$ , a heralding event at  $D_h$  leads to a state with high fidelity to  $|W\rangle_A$  stored in the four ensembles. However, for increasing  $\xi$ ,  $\hat{\rho}_{\geq 2}^{(A)}$  becomes important, leading to modifications of the spin-wave statistics for  $\hat{\rho}_W^{(A)}$  and, thereby, to the entanglement parameters  $\{\Delta, y_c\}$ . Hence, by varying  $\xi$  through the



**Figure 2 | Quadripartite entanglement among four atomic ensembles.**

**a**, Quantum interference between the bipartite entangled pairs of the full quadripartite state (black points) as a function of bipartite phase  $\beta_2$ . **b**, Exploring the entanglement space  $\{\Delta, y_c\}$  for quadripartite states. By controlling the spin-wave statistics, we observe transitions from quadripartite entangled states to tripartite entangled, bipartite entangled and fully separable states (black points). We also display our results for the ‘crossed’ quantum state,  $\hat{\rho}_\times^{(A)}$  (orange points), as further described in Methods. Inset, expanded view of entanglement parameters  $\{\Delta, y_c\}$ . Results for entanglement thermalization, characterized by  $\Delta^{(T)}$  and  $y_c^{(T)}$ , of the spin systems  $\hat{\rho}_G^{(H)}$  and  $\hat{\rho}_G^{(LMG)}$  are shown by the red dashed and blue dash–dot lines, respectively. The red, green and purple bands respectively represent the minimum uncertainties for three-mode ( $\Delta_b^{(3)}$ ) and two-mode entanglement ( $\Delta_b^{(2)}$ ), and for fully separable states ( $\Delta_b^{(1)}$ ); the thickness of each band from the central line corresponds to  $\pm 1$  s.d. of the corresponding bound. In all cases, error bars for the data reflect the statistical and systematic uncertainties (Supplementary Information).

overall intensity for the write beam, we adjust the quantum statistics ( $y_c$ ) and coherence ( $\Delta$ ) of the entangled states  $\hat{\rho}_W^{(A)}$  and  $\hat{\rho}_W^{(Y)}$ .

This procedure is used in Fig. 2b to parametrically increase  $\Delta$  and  $y_c$  in tandem. As  $y_c$  is increased from  $y_c \approx 0$  in the quantum domain to  $y_c \approx 1$  in the classical regime, we observe transitions of the directly measured photonic states  $\hat{\rho}_W^{(Y)}$  (Fig. 2b, black points) from fully quadripartite entangled states ( $\Delta < \Delta_b^{(3)}$ ) to tripartite entangled ( $\Delta_b^{(3)} < \Delta < \Delta_b^{(2)}$ ), to bipartite entangled ( $\Delta_b^{(2)} < \Delta < \Delta_b^{(1)}$ ) and, finally, to fully separable states ( $\Delta_b^{(1)} < \Delta$ ). As shown by the curves, our observations correspond well to a theoretical model of the entanglement parameters,  $\{\Delta^{th}, y_c^{th}\}$ , for entanglement generation, transfer and verification (Supplementary Information). In comparison with our former work on the coherent splitting of a photon<sup>15</sup>, the heralded atomic and photonic W states,  $\hat{\rho}_W^{(A)}$  and  $\hat{\rho}_W^{(Y)}$ , offer qualitatively richer statistical passages through the entanglement spaces delineated by  $\Delta$  and  $y_c$ . Here the quantum coherence ( $\Delta$ ) is intrinsically linked to the statistical character ( $y_c$ ) owing to quantum correlations between the heralding fields,  $\gamma_1$ , and the excitation statistics of the ensembles.

For  $\xi \ll 1$ , the coherent contribution,  $\hat{\rho}_c^{(A)}$ , of the delocalized single quantum strongly dominates any other processes for the full quadripartite state,  $\hat{\rho}_W^{(A)}$ , in equation (3). With a heralding probability  $p_h \approx 3 \times 10^{-4}$  ( $\xi \approx 5 \times 10^{-3}$ ), we achieve the smallest entanglement parameters,  $\Delta^{min} = 0.07 \pm 0.01$  and  $y_c^{min} = 0.038 \pm 0.006$ , for the generated quadripartite entangled states. These parameters are suppressed below the closest three-mode boundary  $\Delta_b^{(3)}$  by ten standard deviations (Supplementary Information). Furthermore, because the local mapping of quantum states from matter to light cannot increase entanglement<sup>7</sup>, our measurements of  $\hat{\rho}_W^{(Y)}$  unambiguously provide a lower bound of the quadripartite entanglement stored in  $\hat{\rho}_W^{(A)}$ . Therefore, the observed strong violation of the uncertainty relations for  $\Delta^{min}$  and  $y_c^{min}$  categorically certifies the creation of measurement-induced entanglement of spin waves among four quantum memories, as well as the coherent transfer of the stored quadripartite entangled states to an entangled state of four propagating electromagnetic fields.

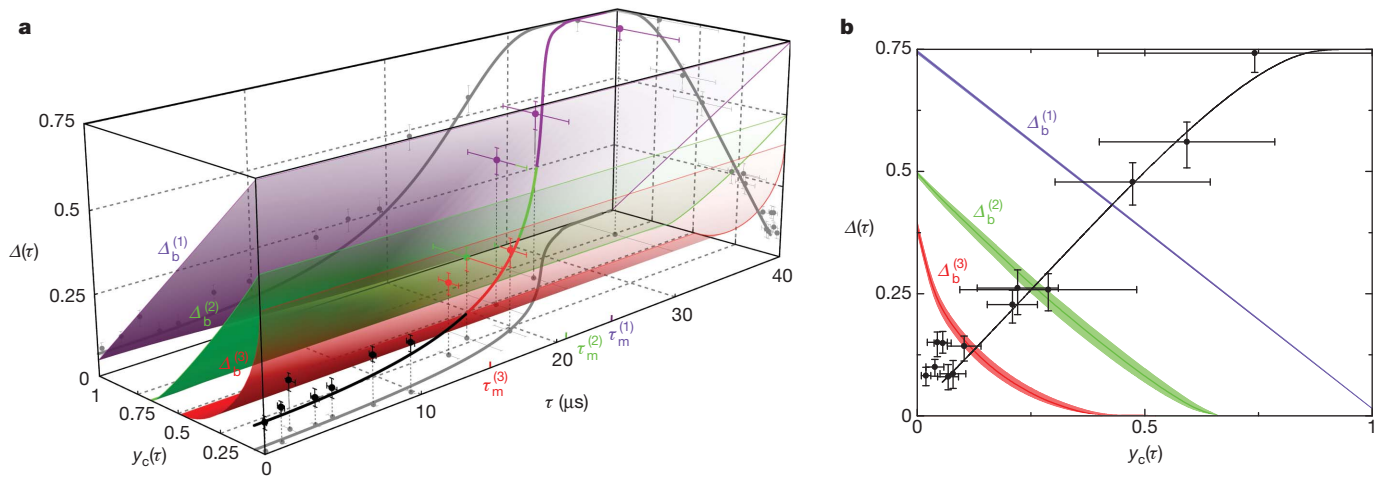
In terms of state fidelity, our approach to the generation of heralded multipartite entanglement compares favourably to matter systems using local interactions (for example trapped ions<sup>23,24</sup>). Despite the intrinsically low preparation probability, the resulting quadripartite entangled state,  $\hat{\rho}_W^{(A)}$ , stored in the four ensembles has high fidelity with the ideal W state, namely  $F^{(A)} = {}_A \langle W | \hat{\rho}_W^{(A)} | W \rangle_A$ . As discussed in Methods, we estimate a lower bound for the unconditional entanglement fidelity of  $F^{(A)} \geq 0.9 \pm 0.1$ , to be compared with the theoretical fidelity,  $F_{th}^{(A)} = 0.98$ , derived for the parameters in our experiment.

Apart from the creation of novel multipartite entangled spin waves, an important benchmark of a quantum interface is the transfer efficiency,  $\lambda$ , of multipartite entanglement from matter to light<sup>11</sup>. Because no known measure applies to our case, we tentatively define the entanglement transfer  $\lambda = F^{(Y)}/F^{(A)}$ , with physical fidelity  $F^{(Y)} = {}_Y \langle W | \hat{\rho}_W^{(Y)} | W \rangle_Y$  for the photonic state  $\hat{\rho}_W^{(Y)}$ . In particular, for  $\xi \ll 1$  we obtain  $F_{th}^{(Y)} \approx \eta_{read} F_{th}^{(A)}$ , which gives  $\lambda_{th} \approx \eta_{read} = 38 \pm 4\%$ , dictated by the retrieval efficiency,  $\eta_{read}$ . Although fidelity is an often used measure, we emphasize that  $F^{(Y)}$  cannot be used to set a threshold for entanglement, because  $\hat{\rho}_W^{(Y)}$  can exhibit multipartite entanglement for any  $F^{(Y)} > 0$ .

To investigate the dynamical behaviour of the observed quadripartite entangled states, we study the temporal evolution of multipartite entanglement stored in the atomic ensembles as a function of a storage time,  $\tau$ . Decoherence for the atomic W state is governed by motional dephasing of spin waves<sup>29</sup>, in which the imprinted atomic phases in  $|\bar{s}_e\rangle$  evolve independently owing to thermal motion, thereby transforming the initial collective state into a subradiant state uncorrelated with the heralding fields,  $\gamma_1$  (Supplementary Information). The net effect is an increase of both entanglement parameters,  $\{\Delta, y_c\}$ , with a timescale  $\tau_m \approx 17 \mu s$  (Methods). Eventually, the growth in  $\Delta(\tau)$  and  $y_c(\tau)$  leads to time-dependent losses of entanglement, marked by successive crossings of the boundaries set by  $\Delta_b^{(3)}$ ,  $\Delta_b^{(2)}$  and  $\Delta_b^{(1)}$ .

In Fig. 3a, we examine the dissipative dynamics of multipartite entanglement for the quantum memories of four ensembles through the evolution of both  $\Delta$  and  $y_c$ . We observe the passage of the initial quadripartite entangled state,  $\hat{\rho}_W^{(A)}(\tau_0)$  at  $\tau_0 = 0.2 \mu s$ , through various domains, progressively evolving from  $M$ -partite entanglement to  $(M-1)$ -partite entanglement at memory times  $\tau = \tau_m^{(M-1)}$ , with the final state,  $\hat{\rho}_W^{(A)}(\tau_f)$ , measured at  $\tau_f = 36.2 \mu s$ . The crossings of the bounds  $\Delta_b^{(3)}$ ,  $\Delta_b^{(2)}$  and  $\Delta_b^{(1)}$  occur at  $\tau_m^{(3)} = 15 \mu s$ ,  $\tau_m^{(2)} = 21 \mu s$  and  $\tau_m^{(1)} = 24 \mu s$ , respectively. In addition, the measured entanglement parameters evolve in qualitative agreement with the simulated dynamics derived for  $\hat{\rho}_W^{(A)}(\tau)$  from our theoretical model (solid line), with deviations (especially for  $\Delta^{th}$ ) discussed in Supplementary Information. Figure 3b shows the parametric losses of entanglement in terms of  $\Delta(\tau)$  and  $y_c(\tau)$ .

Finally, an interesting extension is to relate the characterization of multipartite entanglement by means of  $\{\Delta, y_c\}$  to the relaxations of entanglement in quantum many-body systems<sup>17,18</sup>. We consider two



**Figure 3 | Dissipative dynamics of atomic entanglement.** **a**, Dynamic evolution of entanglement parameters  $\Delta(\tau)$  and  $y_c(\tau)$  for the multipartite quantum state. We observe crossing of the boundaries defined by three-mode (red surface,  $\Delta_b^{(3)}$ ) and two-mode (green surface,  $\Delta_b^{(2)}$ ) entangled states, and separable states (purple surface,  $\Delta_b^{(1)}$ ). We indicate various entanglement orders for quadripartite (black), tripartite (red) and bipartite entangled (green) states, and

fully separable states (purple) for the data points and the curve. The projections of the data points onto the  $y_c$ - $\tau$  and  $\Delta$ - $\tau$  planes show the individual passages of  $\Delta(\tau)$  and  $y_c(\tau)$  (Supplementary Information and Supplementary Movie 1). **b**, Projection of entanglement dynamics onto the  $\Delta$ - $y_c$  plane. The curves in **a** and **b** are from a theoretical model including motional dephasing. Error bars for the data represent the statistical and systematic uncertainties.

ferromagnetic spin models (Heisenberg-like and Lipkin–Meshkov–Glick Hamiltonians  $\hat{H}_H$  and, respectively,  $\hat{H}_{LMG}$ ) as well as their thermal entanglement, as characterized by  $\{\Delta^{(T)}, y_c^{(T)}\}$  (Supplementary Information). Results of our analysis for the Gibbs thermal equilibrium states  $\hat{\rho}_G^{(H)}$  of  $\hat{H}_H$  and  $\hat{\rho}_G^{(LMG)}$  of  $\hat{H}_{LMG}$  are shown by the red dashed and, respectively, blue dash-dot lines in the inset of Fig. 2b. The statistical character of  $\hat{\rho}_W^{(A)}$  for our system of four ensembles follows the thermalization of  $\hat{\rho}_G^{(H)}$  and  $\hat{\rho}_G^{(LMG)}$  for  $y_c \lesssim 0.2$ , whereby  $\hat{\rho}_{\geq 2}^{(A)}$  is thermally populated. This comparison suggests that our method of entanglement characterization could be applied to access the link between off-diagonal long-range order and multipartite entangled spin waves in thermalized quantum magnets<sup>17,18</sup>.

In conclusion, our measurements explicitly demonstrate a coherent matter–light quantum interface for multipartite entanglement by way of the operational metric of quantum uncertainty relations<sup>14–16</sup>. High-fidelity, entangled spin waves are generated in four spatially separated atomic ensembles and coherently transferred to quadripartite entangled beams of light. The quantum memories are individually addressable and can be readily read out at different times for conditional control of entanglement<sup>4</sup>. With recent advances by other groups, the short memory times obtained in Fig. 3 could be improved beyond 1 s (Methods).

Further possibilities include the creation of yet larger multipartite entangled states with efficient scaling<sup>4</sup> for the realization of multipartite quantum networks. For example, quadripartite entangled states of ensemble sets  $\{a, b, c, d\}$  and  $\{a', b', c', d'\}$  could be extended by swapping between  $a$  and  $a'$  to prepare a hexapartite entangled state for  $\{b, b', c, c', d, d'\}$  (Methods). Generalization of such processes will allow the preparation of a single macroscopic entangled state for observing entanglement percolation<sup>6</sup> and extreme non-locality of  $W$  states<sup>30</sup>, as well as for studying quantum phase transitions in strongly correlated systems<sup>17,18</sup>. Finally, the entangled spin waves can be applied to quantum metrology to detect a phase shift of  $\pi$  in an unknown component of  $\hat{\rho}_W^{(A)}$  with efficiency beyond any separable state (Methods).

## METHODS SUMMARY

The preparation stage of our quantum interface lasts  $\Delta t_p = 22$  ms, and consists of laser-cooling and trapping a large cloud of caesium atoms in a magneto-optical trap, from which the atoms are further laser-cooled in an optical molasses and prepared in the state  $|g\rangle$  on release from the trap. We define the four cold atomic ensembles with well-separated optical paths of the quantum fields  $\gamma_1$  and  $\gamma_2$ , which are individually addressed by laser pulses. To operate the quantum interface, we apply a sequence of writing and repumping pulses to the atomic ensembles with a

repetition rate of 2 MHz over  $\Delta t_q = 3$  ms, followed by the next preparation stage. Detection of a spontaneously scattered Raman photon,  $\gamma_1$ , at  $D_h$  triggers a control logic, which terminates the writing and repumping lasers, leaving the ensembles without optical illumination and inhomogeneous broadening, for the quantum storage of heralded multipartite entanglement. The resulting local production rate for the atomic quadripartite entanglement with parameters  $\Delta^{\min}$  and  $y_c^{\min}$  during  $\Delta t_q$  is  $r_q \approx 500$  Hz, giving an average rate of  $r_p \approx 60$  Hz. After a storage time  $\tau$ , read pulses individually transfer the entangled atomic components to propagating multipartite entangled fields,  $\gamma_2$ , via superradiant emissions. In Methods, we describe our spin-wave quantum memories (Fig. 1a, inset) and a control experiment on a ‘crossed’ quantum state,  $\hat{\rho}_\times^{(A)}$ , that results from the intrinsic distinguishability of two bipartite components, as shown in Fig. 2a. We also derive expressions for entanglement fidelity and for the relationship between the set of mutual coherences  $d_{\alpha\beta}$  (between modes  $\alpha, \beta \in \{a_2, b_2, c_2, d_2\}$ ) and  $\Delta$ . In addition, we discuss the prospects for improving our experiment. Finally, we present a quantum-enhanced parameter estimation protocol for using entangled spin waves to detect an atomic phase shift of  $\pi$  with efficiency beyond the limit set by separable states.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 4 July; accepted 5 October 2010.

- Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
- Preskill, J. Quantum computation. *Phys. 219 Course Inf.* (<http://www.theory.caltech.edu/people/preskill/ph219/#lecture>) (1997).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
- Lloyd, S. Universal quantum simulator. *Science* **273**, 1073–1078 (1996).
- Acín, A., Cirac, J. I. & Lewenstein, M. Entanglement percolation in quantum networks. *Nature Phys.* **3**, 256–259 (2007).
- Chou, C. W. *et al.* Measurement-induced entanglement for excitation stored in remote atomic ensembles. *Nature* **438**, 828–832 (2005).
- Moehring, D. L. *et al.* Entanglement of single-atom quantum bits at a distance. *Nature* **449**, 68–71 (2007).
- Simon, J., Tanji, H., Ghosh, S. & Vuletić, V. Single-photon bus connecting spin-wave quantum memories. *Nature Phys.* **3**, 765–769 (2007).
- Weber, B. *et al.* Photon-photon entanglement with a single trapped atom. *Phys. Rev. Lett.* **102**, 030501 (2009).
- Choi, K. S., Deng, H., Laurat, J. & Kimble, H. J. Mapping photonic entanglement into and out of a quantum memory. *Nature* **452**, 67–71 (2008).
- Jost, J. D. *et al.* Entangled mechanical oscillators. *Nature* **459**, 683–685 (2009).
- Sørensen, A. & Mølmer, K. Entanglement and extreme spin squeezing. *Phys. Rev. Lett.* **86**, 4431–4434 (2001).
- Hofmann, H. F. & Takeuchi, S. Violation of local uncertainty relations as a signature of entanglement. *Phys. Rev. A* **68**, 032103 (2003).
- Papp, S. B. *et al.* Characterization of multipartite entanglement for one photon shared among four optical modes. *Science* **324**, 764–768 (2009).



16. Lougovski, P. *et al.* Verifying multipartite mode entanglement of  $W$  states. *N. J. Phys.* **11**, 063029 (2009).
17. Amico, L., Fazio, R., Osterloh, A. & Vedral, V. Entanglement in many-body systems. *Rev. Mod. Phys.* **80**, 517–576 (2008).
18. Gühne, O. & Tóth, G. Entanglement detection. *Phys. Rep.* **474**, 1–75 (2009).
19. Steffen, M. *et al.* Measurement of the entanglement of two superconducting qubits via state tomography. *Science* **313**, 1423–1425 (2006).
20. DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
21. van Enk, S. J., Lütkenhaus, N. & Kimble, H. J. Experimental procedures for entanglement verification. *Phys. Rev. A* **75**, 052318 (2007).
22. Horodecki, R., Horodecki, P., Horodecki, M. & Horodecki, K. Quantum entanglement. *Rev. Mod. Phys.* **81**, 865–942 (2009).
23. Leibfried, D. *et al.* Creation of a six-atom ‘Schrödinger cat’ state. *Nature* **438**, 639–642 (2005).
24. Haffner, H. *et al.* Scalable multiparticle entanglement of trapped ions. *Nature* **438**, 643–646 (2005).
25. Aoki, T. *et al.* Experimental creation of a fully inseparable tripartite continuous-variable state. *Phys. Rev. Lett.* **91**, 080404 (2003).
26. Su, X. *et al.* Experimental preparation of quadripartite cluster and Greenberger-Horne-Zeilinger entangled states for continuous variables. *Phys. Rev. Lett.* **98**, 070502 (2007).
27. Gao, W.-B. *et al.* Experimental demonstration of a hyper-entangled ten-qubit Schrödinger cat state. *Nature Phys.* **6**, 331–335 (2010).
28. Lvovsky, A. I., Sanders, B. C. & Tittel, W. Optical quantum memory. *Nature Photon.* **3**, 706–714 (2009).
29. Simon, J. & Tanji, H. Thompson, J. K. & Vuletić, V. Interfacing collective atomic excitations and single photons. *Phys. Rev. Lett.* **98**, 183601 (2007).
30. Heaney, L., Cabello, A., Santos, M. F. & Vedral, V. Extreme nonlocality with one photon. Preprint at (<http://arxiv.org/abs/0911.0770v2>) (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge discussions with K. Hammerer, P. Zoller and J. Ye. This research is supported by the National Science Foundation, the DOD NSSEFF program, the Northrop Grumman Corporation and the Intelligence Advanced Research Projects Activity. A.G. acknowledges support by the Nakajima Foundation. S.B.P. acknowledges support received as a fellow of the Center for Physics of Information at Caltech.

**Author Contributions** All authors contributed extensively to the research presented in this paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.J.K. ([hjkimble@caltech.edu](mailto:hjkimble@caltech.edu)).

## METHODS

**Experimental details.** The experiment consists of a 22-ms preparation stage and a 3-ms period for operating the quantum interface in Fig. 1 with a repetition rate of 40 Hz and a duty cycle of  $D_c = 3/25$ . For the preparation stage, we load and laser-cool caesium atoms (peak optical depth,  $\sim 30$ ) in a magneto-optical trap for 18 ms, after which the atoms are released from the trap with dynamically compensated eddy currents. The atoms are further cooled in an optical molasses ( $T_f \approx 150$   $\mu$ K) for 3.8 ms and optically pumped to  $|g\rangle$  for 0.2 ms. During this time, a phase reference laser ( $F = 3 \leftrightarrow F' = 4$  transition) also propagates through the atomic ensembles for the active stabilization of the verification interferometer in Fig. 1c by means of *ex situ* phase modulation spectroscopy<sup>15</sup>, which does not affect the operation of the quantum interface (Supplementary Information). Concurrently, dense caesium atoms in paraffin-coated vapour cells located at the heralding and verification ports are prepared in the ground states  $|g\rangle$  and  $|s\rangle$  for filtering the coherent-state lasers scattered into the respective quantum fields,  $\gamma_1$  and  $\gamma_2$ .

**Quantum interface.** For the quantum interface to function during the 3-ms window, in step (i) 20-ns writing pulses (red-detuned by  $\delta = 10$  MHz from the  $|g\rangle \rightarrow |e\rangle$  transition) and 100-ns repumping pulses (resonant with  $|s\rangle \rightarrow |e\rangle$ ) are applied sequentially to the ensembles  $\varepsilon$ , synchronized to a clock running at  $R_c \approx 2$  MHz. This process creates pairwise correlated excitations<sup>4</sup> between the collective atomic modes,  $|\bar{s}_\varepsilon\rangle$ , of the ensembles  $\varepsilon$  and the optical fields  $\gamma_1$  ( $\delta = 10$  MHz below  $|e\rangle \rightarrow |s\rangle$ ). Photodetection of a single photon from the combined fields  $\gamma_1$  at the output of the heralding interferometer effectively erases the ‘which-path’ information for  $\gamma_1$ , and imprints the entangled spin wave  $\hat{\rho}_W^{(A)}$  (equation (3)) onto the ensembles  $a, b, c$  and  $d$  via  $\text{Tr}_h(\hat{\Pi}_h^\dagger \hat{U}_{\text{write}}^\dagger \hat{\rho}_g^{(A)} \hat{U}_{\text{write}})$ . The heralding event at  $D_h$  triggers control logic (Fig. 1a) that deactivates intensity modulators of the writing (IM<sub>write</sub>), repumping and reading lasers (IM<sub>read</sub>) for the quantum storage of  $\hat{\rho}_W^{(A)}$  in step (ii). After a user-controlled delay,  $\tau$ , step (iii) is initiated with 20-ns, strong read pulses (Rabi frequency of 24 MHz, resonant with  $|s\rangle \rightarrow |e\rangle$ ) that address the ensembles in Fig. 1c and coherently transfer the entangled atomic components  $a, b, c$  and  $d$  of  $\hat{\rho}_W^{(A)}$  ( $\tau$ ) one by one to propagating beams  $\gamma_2 \in \{a_2, b_2, c_2, d_2\}$  (resonant with  $|e\rangle \rightarrow |g\rangle$ ), comprising the entangled photonic state  $\hat{\rho}_W^{(v)}$  ( $\tau$ ), via the operation  $\hat{\rho}_W^{(v)} = \text{Tr}_A(\hat{U}_{\text{read}}^\dagger \hat{\rho}_W^{(A)} \hat{U}_{\text{read}})$ . Here  $\text{Tr}_A$  traces over the atomic systems that are later shelved into the ground states  $|\bar{g}_\varepsilon\rangle$ . The retrieval efficiency,  $\eta_{\text{read}}$ , is collectively enhanced for large  $N_A$  (ref. 4), leading to  $\eta_{\text{read}} = 0.38 \pm 0.06$  in our experiment. The average production rate for the atomic quadripartite entanglement for  $\{\Delta^{\text{min}}, y_c^{\text{min}}\}$  is  $r_p = R_c D_c p_h \approx 60$  Hz, and the actual rate during the 3-ms operating window is  $r_q = R_c p_h \approx 500$  Hz. The atomic level diagrams for entanglement generation and quantum state exchanges are shown as insets to Fig. 1b, c. States  $|g\rangle$  and  $|s\rangle$  are the hyperfine ground states  $F = 4$  and  $F = 3$  of  $6S_{1/2}$  in atomic caesium, respectively; state  $|e\rangle$  is the hyperfine level  $F' = 4$  of the electronic excited state  $6P_{3/2}$ .

**Spin-wave quantum memories.** The quantum information of the entangled state for equation (1) is encoded in the quantum numbers of spin waves (collective excitations) for the pseudo-spin of the hyperfine ground electronic levels  $6S_{1/2}$  ( $F = 3, F = 4$ ) in atomic caesium. The fluorescence images shown in the inset of Fig. 1a depict the collective atomic modes of ensembles  $\varepsilon \in \{a, b, c, d\}$  for exciting the entangled spin waves  $\hat{\rho}_W^{(A)}$  with 1-mm separations and 60- $\mu$ m waists. The geometry of the collective excitations for the four ensembles  $a, b, c$  and  $d$  is defined by the point-spread functions of the imaging systems for the fields  $\gamma_1$  and  $\gamma_2$ , where each ensemble consists of a cold cloud of  $N_{A,\varepsilon} \approx 10^6$  caesium atoms. We use an off-axis configuration<sup>31</sup> to address each ensemble  $\varepsilon$  individually, with an angle of  $\theta = 2.5^\circ$  between the classical and non-classical beams (Supplementary Information), that creates spin waves  $|\bar{s}_\varepsilon\rangle$  associated with wavevectors  $\delta \mathbf{k} = \mathbf{k}_{\text{write}} - \mathbf{k}_{\gamma_1}$  for each  $\varepsilon$ . These spin waves are analogous to other types of collective excitation in many-body systems, such as magnons and plasmons, and can be converted to dark-state polaritons for the coherent transfer,  $\hat{U}_{\text{read}}$ , of entanglement. For the phase-matching configuration and temperature of our ensembles, the memory times  $\tau_m^{(3)}$ ,  $\tau_m^{(2)}$  and  $\tau_m^{(1)}$  (Fig. 3) are dominantly determined by the motional dephasing of the spin waves  $|\bar{s}_\varepsilon\rangle$  (ref. 29). For a thermal velocity of  $v_t \approx 14$  cm s<sup>-1</sup>, we estimate a memory time of  $\tau_m \approx (0.85 \mu\text{m}) / 4\pi \sin(\theta/2) v_t = 17$   $\mu$ s. However, the ground-state dephasing due to inhomogeneous broadening is expected to be  $> 50$   $\mu$ s in our experiment, as inferred from two-photon Raman spectroscopy.

**Quantum uncertainty relations and genuine multipartite entanglement.** To verify the entanglement by way of  $\Delta$  and  $y_c$ , we first evaluate the photon statistics  $p_0, p_1$  and  $p_{\geq 2}$  for the measurement of  $y_c$ . Operationally, this is accomplished by measuring the individual probabilities,  $q_{ijkl}$  for  $i, j, k, l \in \{0, 1\}$  photons to occupy the respective optical modes  $\gamma_2 \in \{a_2, b_2, c_2, d_2\}$  at the output faces of the ensembles, through photoelectric detections  $\hat{I}_i^{(s)}$ . For the measurement of  $\Delta$ , we quantify the off-diagonal coherence,  $\bar{d}$ , of  $\hat{\rho}_W^{(v)}$  by pairwise interferences of all possible sets of modes  $\alpha, \beta \in \{a_2, b_2, c_2, d_2\}$  with the verification interferometer. The photon probabilities  $p_{1000}, p_{0100}, p_{0010}$  and  $p_{0001}$  at the output modes of the verification interferometer thereby result from the coherent interferences of the four purportedly

entangled fields  $\gamma_2$  that depend on the phase orientations  $\{\beta_1, \beta_2, \beta_3\}_v$  of  $\hat{I}_i^{(c)}$  (Supplementary Information).

Our conclusion of genuine multipartite entanglement for the atomic and photonic states  $\{\hat{\rho}_W^{(A)}, \hat{\rho}_W^{(v)}\}$  does not rely on weaker conditions based on the non-separability along any fixed bipartition of  $\{\hat{\rho}_W^{(A)}, \hat{\rho}_W^{(v)}\}$ . The genuine  $M$ -partite entangled states created from our experiment can only be represented as mixtures of pure states that all possess  $M$ -partite entanglement, as for the case of genuine ‘ $k$ -producibility’ in multipartite spin models<sup>17,18</sup>. We note that our entanglement verification protocol cannot be used to verify the absence of entanglement for the physical state  $\hat{\rho}_W^{(v)}$  in an infinite dimension<sup>32</sup>. Finally, we emphasize that our analysis makes use of the full physical state,  $\{\hat{\rho}_W^{(A)}, \hat{\rho}_W^{(v)}\}$ , including the vacuum component,  $\hat{\rho}_0$ , and higher-order terms,  $\hat{\rho}_{\geq 2}$ , and does not rely upon a spurious postdiction based on a preferred set of detection events (Supplementary Information).

**Generation and characterization of a ‘crossed’ quantum state.** As a control experiment, we reconfigure the heralding interferometer such that path information could in principle be revealed up to the bipartite split of the ensemble pairs  $\{a, b\}$  and  $\{c, d\}$  by analysing the polarization state of the heralding photon,  $\gamma_1$ . In this case, the heralding measurement,  $\hat{I}_\times$ , prepares a ‘crossed’ atomic state,  $\hat{\rho}_\times^{(A)}$ , with no coherence shared between  $\{a, b\}$  and  $\{c, d\}$ . Thus, we observe an absence of interference in Fig. 2a (orange points). However, this modified  $\hat{I}_\times$  preserves the bipartite entanglement within  $\{a, b\}$  and  $\{c, d\}$ , which explains our observation of the uncertainty,  $\Delta$ , reduced below the one-mode bound,  $\Delta_b^{(1)}$ , for  $y_c = 0.07 \pm 0.01$ . Similarly, we also detect the statistical transition from bipartite entanglement to fully separable states for the ‘crossed’ state in Fig. 2b, despite the disentanglement for the bipartition  $(|)\rangle$  of  $\{a, b\}|\{c, d\}$ .

**Relationship between quantum uncertainty and off-diagonal coherences.** Here we derive the general expression for the upper bound of the sum uncertainty,  $\Delta$ , as a function of the coherence,  $\bar{d}$ . First we note that  $\Delta$  is only sensitive to the one-excitation subspace,  $\hat{\rho}_1$ , of  $\hat{\rho}_r$

$$\hat{\rho}_1 = \begin{pmatrix} s_{1000} & d_{ab} & d_{ac} & d_{ad} \\ d_{ba}^* & s_{0100} & d_{bc} & d_{bd} \\ d_{ca}^* & d_{cb}^* & s_{0010} & d_{cd} \\ d_{da}^* & d_{db}^* & d_{dc}^* & s_{0001} \end{pmatrix}$$

normalized such that  $\text{Tr}(\hat{\rho}_1) = s_{1000} + s_{0100} + s_{0010} + s_{0001} = 1$ . Here the diagonal elements,  $\mathbf{s}_1 = (s_{1000}, s_{0100}, s_{0010}, s_{0001})$ , of  $\hat{\rho}_1$  are related to the one-photon probabilities,  $\mathbf{q}_1 = (q_{1000}, q_{0100}, q_{0010}, q_{0001})$ , at the faces of the ensembles by  $p_1 \mathbf{s}_1 = \mathbf{q}_1$ . By transforming  $\hat{\rho}_1$  into the basis spanned by  $|W_i\rangle_v$ , we then find the expressions for the normalized output photon probabilities,  $p_{1000}, p_{0100}, p_{0010}$  and  $p_{0001}$ , of the verification interferometer as functions of  $\mathbf{s}_1$  and  $d_{\alpha\beta}$ . The sum uncertainty,  $\Delta$ , is then expressed as

$$\Delta = \frac{3}{4} - \{(|d_{ab}| + |d_{cd}|)^2 + (|d_{ac}| + |d_{bd}|)^2 + (|d_{ad}| + |d_{bc}|)^2\}$$

Thus, we obtain  $\Delta \lesssim (3/4)(1 - 16\bar{d}^2)$ . The average value of the six unique off-diagonal elements is  $\bar{d} = (1/6) \sum_{\alpha, \beta} |d_{\alpha\beta}|$  with  $0 \leq \bar{d} \leq 1/4$ , and the effective interference visibility is given by  $V_{\text{eff}} = 4\bar{d}$ .

**Derivation of entanglement fidelity.** Here we obtain the expression for the lower-bound unconditional entanglement fidelity,  $F^{(A)} = \tilde{p}_1 F_1$ , where  $\tilde{p}_1$  is the probability of a single spin wave,  $\hat{\rho}_1^{(A)}$ , in the heralded state  $\hat{\rho}_W^{(A)}$  and  $F_1 = \langle W_1 | \hat{\rho}_1^{(A)} | W_1 \rangle$  is the conditional fidelity for  $\hat{\rho}_1^{(A)}$ . We start by noting that the projective measurement  $\hat{I}_i^{(c)}$  for  $\Delta$  gives the conditional fidelity,  $F_1$ , of  $\hat{\rho}_r$  projected onto one of four orthonormal  $W$  states,  $|W_i\rangle_v = |W_1\rangle_v$ , for example  $|1000\rangle + e^{i\beta_1}|0100\rangle + e^{i\beta_2}|0010\rangle + e^{i\beta_3}|0001\rangle$ . Hence, we can define  $\Delta = 1 - F_1^2 - \sum_{i=2}^4 F_i^2$  in terms of the respective overlaps  $F_i$ . Because of the orthonormality condition,  $\sum_{i=1}^4 F_i = 1$ , the sum uncertainty is bounded by  $\Delta \geq 1 - F_1^2 - (1 - F_1)^2$ , from which we obtain  $F_1 \geq \sqrt{(1/2)(1/2 - \Delta)} + 1/2$ . Finally, by multiplying a factor of  $\tilde{p}_1$ , the probability of exciting one spin wave distributed among the four ensembles, we find the lower-bound fidelity  $F^{(A)} \geq \tilde{p}_1(\sqrt{(1/2)(1/2 - \Delta)} + 1/2)$  obtained unconditionally for the heralded atomic state  $\hat{\rho}_W^{(A)}$ . In principle, the imbalances in the interferometer can rotate the projectors into non-orthonormal sets<sup>16</sup>. However, the measured losses and the beam-splitter ratios are well-enough balanced that any changes in  $F^{(A)}$  due to modified projectors are well within the uncertainties of the data, as evidenced by the close-to-unity projection fidelity,  $F^{(r)} = 99.9^{+0.1}_{-0.2}\%$  (Supplementary Information). In the experiment,  $\tilde{p}_1$  and  $F_1$  are determined from the inferences of the spin-wave statistics (by means of  $y_c$ ) and the coherences (by means of  $\Delta$ ), respectively.

**Prospects for improving memory time and matter–light transfer efficiency.** By operating the clock speed at  $R_c \rightarrow 10$  MHz and  $\tau_m^{(3)} \approx 20$   $\mu$ s, we could prepare hexapartite ( $M = 6$ ) entanglement with probability  $3\tau_m \eta_{\text{read}} p_h^2 / 8 \approx 10^{-5}$  by connecting two quadripartite states  $\hat{\rho}_W^{(A)}$  for  $\Delta^{\text{min}}$  and  $y_c^{\text{min}}$ , with enhancement

factor  $z = 400$  (ref. 33), thereby giving a local production rate of  $r_q \approx 50\text{--}100\text{ Hz}$ , or an average rate of  $r_p \approx 5\text{--}10\text{ Hz}$  with our current duty cycle,  $D_c$ . The most challenging aspect of verifying the hexapartite entangled states is the quantification of the higher-order contamination,  $\hat{\rho}_{\geq 2}$ , which we estimate to be one event per 10 h. This integration rate is feasible with our current system. More generally,  $M_1$ -partite and  $M_2$ -partite entangled states can be fused by entanglement connection to create an  $M = (M_1 + M_2 - 2)$ -partite entangled state. However, the memory times  $\tau_m^{(3)}$ ,  $\tau_m^{(2)}$  and  $\tau_m^{(1)}$  (Fig. 3) and the entanglement transfer,  $\lambda$ , from matter to light limit our ability to scale the multipartite entanglement beyond  $M > 6$  by way of conditional control and connection of entanglement<sup>33,34</sup> with our current experimental parameters.

The prerequisite storage techniques for suppressing both the internal and the motional spin-wave dephasings can be extended for  $\tau_m$  with advances in ensemble-based quantum memories<sup>35,36,37</sup>). Recent experiments with single ensembles have achieved coherence times of up to  $\tau_m \approx 1.5\text{ s}$  in quantum degenerate gases<sup>38,39</sup>, albeit with efficiencies of  $\lesssim 1\%$ . Also, the transfer efficiency can be increased to  $\lambda_{\text{th}} \approx 0.9$  by enclosing the ensembles within high-finesse cavities<sup>29</sup>. System integrations by way of atom-chip technology and waveguide coupling<sup>40,41</sup> hold great potential for scalability given the strong cooperativity and the long coherence<sup>42</sup>. At this level, two or more heralded processes of multipartite entanglement generation can be made on-demand on timescales of  $\tau_{\text{det}} \approx 1/R_{\text{cph}} = 1\text{ ms}$ , with  $\tau_m \gg \tau_{\text{det}}$  (refs 33, 34).

Realistically, the expansion of multipartite entangled states  $\hat{\rho}_W^{(A)}$  will be limited by the intrinsic degradations of the entanglement parameters  $\Delta$  and  $\gamma_c$ , which inevitably increase with each step of entanglement connection<sup>34</sup>, and by the specific quantum repeater architecture implemented on  $\hat{\rho}_W^{(A)}$ . The latter is an extremely rich area of research in view of the large classes of methods for connecting multipartite entangled states, making it premature to specify a particular architecture for multipartite entanglement expansion. However, our experiment will hopefully stimulate theoretical studies of complex repeater architectures for multipartite systems, beyond traditional one-to-one networks<sup>43</sup>.

**Quantum-enhanced parameter estimation with entangled spin waves.** We describe a quantum-enhanced parameter estimation protocol whereby a phase shift in a single ensemble,  $e_i$ , of the quadripartite state  $e_i \in \{a, b, c, d\}$  can be detected with efficiency beyond that for any separable state. Specifically, we consider a  $\pi$  phase shift,  $\hat{U}_{\pi, e_i} = \exp(i\pi \hat{n}_{e_i})$ , applied to an unknown spin-wave component  $e_i \in \{a, b, c, d\}$  ( $\hat{n}_{e_i} = \hat{S}_{e_i}^z \hat{S}_{e_i}^z$ ) of the atomic state  $\hat{\rho}_W^{(A)}$ , or to a spatial field mode,  $\gamma_{2i} \in \{\gamma_{2a}, \gamma_{2b}, \gamma_{2c}, \gamma_{2d}\}$ , of the photonic state  $\hat{\rho}_W^{(\gamma)}$  ( $\hat{n}_{\gamma_{2i}} = \hat{a}_{\gamma_{2i}}^\dagger \hat{a}_{\gamma_{2i}}$ ). Our goal is to find the  $\pi$ -phase-shifted ensemble,  $e_i$  (or optical mode,  $\gamma_{2i}$ ), in a single measurement under the condition that an average of one spin wave is populated in total; that is,  $\sum_i \text{Tr}(\hat{n}_{e_i} \hat{\rho}_W^{(A)}) = 1$  (or, for optical modes,  $\sum_i \text{Tr}(\hat{n}_{\gamma_{2i}} \hat{\rho}_W^{(\gamma)}) = 1$ ). As a quantum benchmark, we consider an average success probability  $P_s = (1/4) \sum_{e_i} \text{Tr}(\hat{U}_{\pi, e_i}^{(u)} \hat{U}_{\pi, e_i}^{(u)} \hat{\rho}_W^{(A)} \hat{U}_{\pi, e_i})$  (failure probability,  $P_f = 1 - P_s$ ) for distinguishing the phase-shifted ensemble  $e_i$  (or mode  $\gamma_{2i}$ ) among the four possibilities  $e_i \in \{a, b, c, d\}$  (or  $\gamma_{2i} \in \{\gamma_{2a}, \gamma_{2b}, \gamma_{2c}, \gamma_{2d}\}$ ) by way of unambiguous quantum state discrimination,  $\hat{\Pi}_{e_i}^{(u)}$  (refs 44–47).

First we consider an ideal W state,  $|W\rangle_o = |W\rangle_A$  (or  $|W\rangle_{\gamma_2}$ ), with atomic phases  $\phi_i \in \{\phi_1, \phi_2, \phi_3\}$  (or photonic phases  $\phi'_i \in \{\phi'_1, \phi'_2, \phi'_3\}$ ). In this case, the  $\pi$ -phase-shifted entangled W states  $|W_{e_i}\rangle_f \in \{|W_a^{(\pi)}\rangle_f, |W_b^{(\pi)}\rangle_f, |W_c^{(\pi)}\rangle_f, |W_d^{(\pi)}\rangle_f\}$  can be detected deterministically, because  $|W_{e_i}^{(\pi)}\rangle_f = \hat{U}_{\pi, e_i} |W\rangle_o$  forms an orthonormal complete set that spans the state space of  $\hat{\rho}_1$ , resulting from the underlying symmetry of  $|W\rangle_o$  with respect to any rotation  $\hat{U}_{\pi, e_i}$  on a generalized Bloch sphere. Operationally, we set the verification phases  $\beta_{1,2} - \phi'_{1,2} = 0$  and  $\beta_3 - \phi'_3 = \pi$ . Then the  $\pi$ -phase-shifted ensemble,  $e_i$ , can be unambiguously distinguished because the otherwise balanced output photon probabilities,  $\mathbf{p}_v = (p_{1000}, p_{0100}, p_{0010}, p_{0001}) = (0.25, 0.25, 0.25, 0.25)$ , of the verification interferometer will be transformed to  $\mathbf{p}_v = (1, 0, 0, 0)$ , for a  $\pi$  phase induced in ensemble  $a$ , to  $\mathbf{p}_v = (0, 1, 0, 0)$  in ensemble  $b$ , to  $\mathbf{p}_v = (0, 0, 1, 0)$  in ensemble  $c$  and to  $\mathbf{p}_v = (0, 0, 0, 1)$  in ensemble  $d$ , each with success probability  $P_s^{(\text{ent})} = 1$ .

For fully separable states  $|\Psi\rangle_o = |\psi_a\rangle_a |\psi_b\rangle_b |\psi_c\rangle_c |\psi_d\rangle_d$ , with  $|\psi_{e_i}\rangle_{e_i} = \sum_{n=0}^{\infty} c_{e_i}^{(n)} |n\rangle_{e_i}$ , we displace the resulting  $\pi$ -phase-shifted state,  $|\Psi_{e_i}^{(\pi)}\rangle_f = \hat{U}_{\pi, e_i} |\Psi\rangle_o$ , with a local unitary transformation,  $\hat{V}_{e_i} |\psi_{e_i}\rangle_{e_i} = |0\rangle_{e_i}$ . The overall process,  $\hat{V}_a \hat{V}_b \hat{V}_c \hat{V}_d \hat{U}_{\pi, e_i}$ , maps the initial product state,  $|\Psi\rangle_o$ , into

$\hat{V}_a \hat{U}_{\pi, a} |\psi_a\rangle_a |0\rangle_b |0\rangle_c |0\rangle_d$  (phase shift in ensemble  $a$ ),  $|0\rangle_a \hat{V}_b \hat{U}_{\pi, b} |\psi_b\rangle_b |0\rangle_c |0\rangle_d$  (ensemble  $b$ ),  $|0\rangle_a |0\rangle_b \hat{V}_c \hat{U}_{\pi, c} |\psi_c\rangle_c |0\rangle_d$  (ensemble  $c$ ) or  $|0\rangle_a |0\rangle_b |0\rangle_c \hat{V}_d \hat{U}_{\pi, d} |\psi_d\rangle_d$  (ensemble  $d$ ), with only one  $e_i$  containing  $\langle \hat{n}_{e_i} \rangle > 0$  excitations. Thus, we can unambiguously identify the phase-shifted ensemble given a photodetection, albeit with a failure probability

$$P_f = \frac{1}{4} \sum_{e_i} \left| \langle 0 | \hat{V}_{e_i} \hat{U}_{\pi, e_i} |\psi_{e_i}\rangle_{e_i} \right|^2 \\ = \frac{1}{4} \sum_{e_i} \left| \langle \psi_{e_i} | \hat{U}_{\pi, e_i} |\psi_{e_i}\rangle_{e_i} \right|^2$$

arising from inconclusive null events (that is,  $|0000\rangle\langle 0000|$ ). We derive the maximum success probability,  $P_s^{(\text{max})} = 1 - P_f^{(\text{min})}$ , and the optimal state,  $|\Psi\rangle_o = |\Psi\rangle_{\text{opt}}$ , by minimizing  $P_f^{(\text{min})}$  over all possible realizations of  $c_{e_i}^{(n)}$  satisfying  $\sum_{e_i} \langle \psi_{e_i} | \hat{n}_{e_i} | \psi_{e_i} \rangle_{e_i} = 1$ . Specifically, we find that an optimal (pure) separable state  $|\Psi\rangle_{\text{opt}} = \Pi_{e_i} \left( \sqrt{3/4} |0\rangle_{e_i} + \sqrt{1/4} |1\rangle_{e_i} \right)$  can be used for the parameter estimation protocol to infer  $e_i$  with  $P_s^{(\text{max})} = 0.75$ . Similarly, the maximum success probability  $P_s^{(\text{coh})}$  can be derived for multimode coherent states  $\Pi_{e_i} |\alpha_{e_i}\rangle_{e_i}$ , giving a classical bound of  $P_s^{(\text{coh})} = 1 - 1/e$ .

Finally, we consider the upper bound  $P_s^{(\text{max})}$  for mixed separable states,  $\hat{\rho}_o^{(\text{sep})}$ , with pure state decompositions  $\hat{\rho}_o^{(\text{sep})} = \sum_m p_m |\Psi_m\rangle_o \langle \Psi_m|$ . Generally, the transformations  $\hat{V}_{e_i}$ , as discussed above, do not exist for  $\hat{\rho}_o^{(\text{sep})}$ , precluding the possibility of unambiguous state discrimination. Thus, the success probability,  $P_s(\hat{\rho}_o^{(\text{sep})})$ , is bounded from above by the convex combinations of  $|\Psi_m\rangle_o$ , such that

$$P_s(\hat{\rho}_o^{(\text{sep})}) \leq \sum_m p_m P_s(|\Psi_m\rangle_o \langle \Psi_m|) \leq P_s^{(\text{max})} = 0.75$$

Importantly, the maximum success probability,  $P_s^{(\text{max})} = 0.75$ , attainable for any  $\hat{\rho}_o^{(\text{sep})}$ , is less than  $P_s^{(\text{ent})} = 1$  for entangled states  $|W\rangle_o$ . Thus, the entangled spin waves in the experiment can be used to sense an atomic phase shift beyond the limit for any unentangled state. A comprehensive analysis of our protocol, including experimental imperfections (for example detection efficiency) and other measurement strategies, will be discussed elsewhere.

- Balić, V., Braje, D. A., Kolchin, P., Yin, G. Y. & Harris, S. E. Generation of paired photons with controllable waveforms. *Phys. Rev. Lett.* **94**, 183601 (2005).
- Eisert, J., Simon, C. & Plenio, M. B. On the quantification of entanglement in infinite-dimensional quantum systems. *J. Phys. A* **35**, 3911–3923 (2002).
- Chou, C.-W. *et al.* Functional quantum nodes for entanglement distribution over scalable quantum networks. *Science* **316**, 1316–1320 (2007).
- Laurat, J. *et al.* Towards experimental entanglement connection with atomic ensembles in the single excitation regime. *N. J. Phys.* **9**, 207–220 (2007).
- Hammerer, K., Sørensen, A. S. & Polzik, E. S. Quantum interface between light and atomic ensembles. *Rev. Mod. Phys.* **82**, 1041–1093 (2010).
- Zhao, R. *et al.* Long-lived quantum memory. *Nature Phys.* **5**, 100–104 (2009).
- Zhao, B. *et al.* A millisecond quantum memory for scalable quantum networks. *Nature Phys.* **5**, 95–99 (2009).
- Schnorrberger, U. *et al.* Electromagnetically induced transparency and light storage in an atomic Mott insulator. *Phys. Rev. Lett.* **103**, 033003 (2009).
- Zhang, R., Garner, S. R. & Hau, L. V. Creation of long-term coherent optical memory via controlled nonlinear interactions in Bose-Einstein condensates. *Phys. Rev. Lett.* **103**, 233602 (2009).
- Colombe, Y. *et al.* Strong atom-field coupling for Bose-Einstein condensates in an optical cavity on a chip. *Nature* **450**, 272–276 (2007).
- Vetsch, E. *et al.* Optical interface created by laser-cooled atoms trapped in the evanescent field surrounding an optical nanofiber. *Phys. Rev. Lett.* **104**, 203603 (2010).
- Deutsch, C. *et al.* Spin self-rephasing and very long coherence times in a trapped atomic ensemble. *Phys. Rev. Lett.* **105**, 020401 (2010).
- Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. Preprint at (<http://arxiv.org/abs/0906.2699v2>) (2009).
- Ivanovic, I. D. How to differentiate between non-orthogonal states. *Phys. Lett. A* **123**, 257–259 (1987).
- Dieks, D. Overlap and distinguishability of quantum states. *Phys. Lett. A* **126**, 303–306 (1988).
- Peres, A. How to differentiate between non-orthogonal states. *Phys. Lett. A* **128**, 19 (1988).
- Cheffes, A. Unambiguous discrimination between linearly independent quantum states. *Phys. Lett. A* **239**, 339–347 (1998).



# Free-standing mesoporous silica films with tunable chiral nematic structures

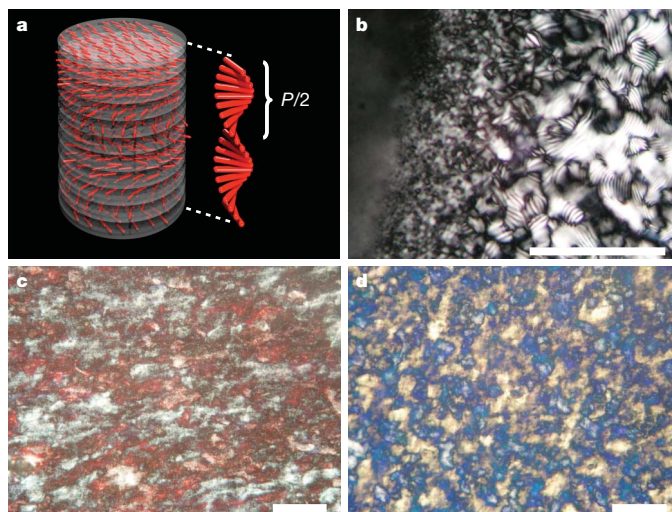
Kevin E. Shopsowitz<sup>1</sup>, Hao Qi<sup>1</sup>, Wadood Y. Hamad<sup>2</sup> & Mark J. MacLachlan<sup>1</sup>

Chirality at the molecular level is found in diverse biological structures, such as polysaccharides, proteins and DNA, and is responsible for many of their unique properties<sup>1</sup>. Introducing chirality into porous inorganic solids may produce new types of materials that could be useful for chiral separation, stereospecific catalysis, chiral recognition (sensing) and photonic materials<sup>2–5</sup>. Template synthesis of inorganic solids using the self-assembly of lyotropic liquid crystals offers access to materials with well-defined porous structures<sup>6–12</sup>, but only recently has chirality been introduced into hexagonal mesostructures through the use of a chiral surfactant<sup>13,14</sup>. Efforts to impart chirality at a larger length scale using self-assembly are almost unknown. Here we describe the development of a photonic mesoporous inorganic solid that is a cast of a chiral nematic liquid crystal formed from nanocrystalline cellulose. These materials may be obtained as free-standing films with high surface area. The peak reflected wavelength of the films can be varied across the entire visible spectrum and into the near-infrared through simple changes in the synthetic conditions. To the best of our knowledge these are the first materials to combine mesoporosity with long-range chiral ordering that produces photonic properties. Our findings could lead to the development of new materials for applications in, for example, tuneable reflective filters and sensors. In addition, this type of material could be used as a hard template to generate other new materials with chiral nematic structures.

The chiral nematic (or cholesteric) liquid crystalline phase, in which mesogens are organized in a helical assembly, was first observed for cholesterol derivatives but is now known for a variety of molecules and polymers. The helical organization of a chiral nematic liquid crystal causes angle-dependent selective reflection of circularly polarized light, which results in iridescence when the helical pitch is of the order of the wavelength of visible light. For this reason, chiral nematic liquid crystals have been extensively studied for their photonic properties and used for applications such as polarizing mirrors, reflective displays and lasers<sup>15–17</sup>. Chiral nematic phases have also been exploited for other applications such as the synthesis of helical polymers<sup>18</sup>. In nature, the solid-state chiral nematic organization of chitin results in the brilliant iridescent colours of beetle exoskeletons<sup>19</sup>.

Stable nanocrystals of cellulose may be obtained by acid-catalysed hydrolysis of bulk cellulose<sup>20</sup>. In water, suspensions of nanocrystalline cellulose (NCC) organize into a chiral nematic phase (Fig. 1a) that can be preserved upon air-drying, resulting in iridescent films<sup>21,22</sup>. These properties, along with the high surface area of NCC, make it an interesting potential template for porous inorganic materials. Researchers have previously attempted to use the chiral nematic phase to template mesoporous solids. Mann and co-workers showed that NCC can be used to template birefringent silica, but no long-range helical ordering was observed and no porosity was measured<sup>23</sup>. Using the chiral nematic phase of hydroxypropylcellulose as a template, Thomas and Antonietti obtained high-surface area porous silica<sup>24</sup>. Chiral nematic organization, however, did not appear to be retained in the pure silica replicas.

In our experiments, we used NCC prepared by sulphuric acid hydrolysis of bleached kraft softwood pulp as a chiral nematic template for mesoporous silica. Aqueous suspensions of NCC (3 wt%) were typically used, at pH = 2.4 before the addition of the silica precursor. We found that at this pH, Si(OEt)<sub>4</sub> (TEOS) or Si(OMe)<sub>4</sub> (TMOS) could be hydrolysed in the presence of NCC to give a homogeneous mixture without disrupting the ability of NCC to form a chiral nematic phase. Polarized optical microscopy (POM) showed the formation of a fingerprint texture during evaporation, indicating that the chiral nematic phase is established during drying in the presence of the silica precursor (Fig. 1b). After drying, free-standing composite films were obtained. Visually, as well as by POM (Fig. 1c) and scanning electron microscopy (SEM) (Supplementary Fig. 6), the free-standing composite films are very similar to those composed of pure NCC. However, in contrast to pure NCC films, the composite films cannot readily be suspended in water. Circular dichroism demonstrated that the composite films have left-handed helical structures, which is also observed for pure NCC films<sup>25</sup>. The pH range over which this procedure is successful was found to be narrow. Attempts to adjust the pH of the NCC suspension before the synthesis all resulted in a disruption of chiral nematic ordering in the composite films. However, at pH = 2.4



**Figure 1 | Schematic of the chiral nematic ordering of NCC crystallites and POM images.** **a**, Schematic of the chiral nematic ordering present in NCC, along with an illustration of the half-helical pitch  $P/2$  (~150–650 nm). **b**, POM image of a TEOS/NCC suspension observed during slow evaporation at room temperature (22 °C) clearly shows a fingerprint texture characteristic of chiral nematic ordering. **c**, POM image of an NCC/silica composite film. Strong birefringence and domains with different orientations are present. **d**, POM image of the mesoporous silica film obtained from the calcination of the film in **c**. A shift in colour from red to blue was observed, while the overall texture remained essentially unchanged. All micrographs were taken with crossed polarizers (scale bar, 100 µm).

<sup>1</sup>Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia, V6T 1Z1, Canada. <sup>2</sup>FPIInnovations, 3800 Westbrook Mall, Vancouver, British Columbia, V6S 2L9, Canada.

the synthesis was quite robust with respect to the amount of silica precursor used. This allowed for the synthesis of chiral nematic films with a wide range of compositions (30–70 wt% silica by thermal gravimetric analysis).

The peak wavelength reflected by a chiral nematic structure ( $\lambda_{\text{max}}$ ) for incident light normal to the surface may be expressed as:

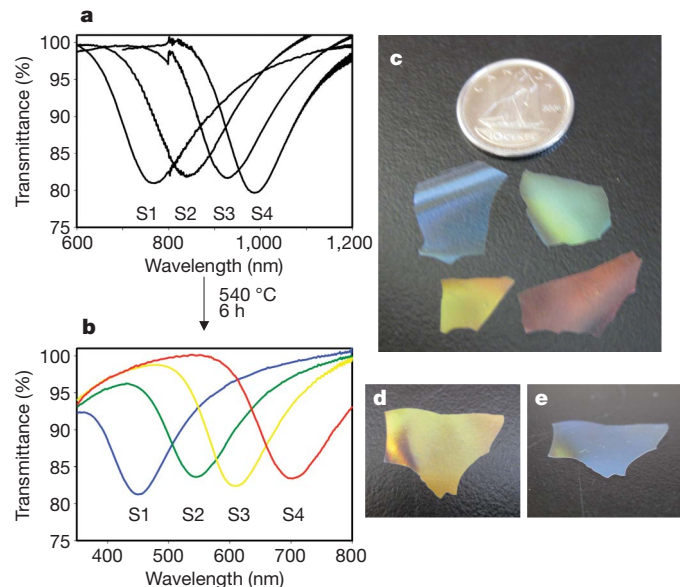
$$\lambda_{\text{max}} = n_{\text{avg}} P \quad (1)$$

where  $n_{\text{avg}}$  is the average refractive index and  $P$  is the helical pitch<sup>26</sup>.  $\lambda_{\text{max}}$  may therefore be tuned by altering the helical pitch or the average refractive index of a chiral nematic material. We were able to vary  $\lambda_{\text{max}}$  of the composite films from the visible to the near-infrared by increasing the proportion of silica precursor relative to NCC (Fig. 2a). The average refractive indices of the different composite materials are essentially constant because the two components, SiO<sub>2</sub> and crystalline cellulose, have similar refractive indices ( $n = 1.46$  and  $1.54$ , respectively). The observed increase in  $\lambda_{\text{max}}$  for samples with higher silica content must therefore be the result of an increase in  $P$ . The increase in helical pitch may be caused by greater silica wall thickness as well as repulsive interactions between the negatively charged silica species and cellulose nanocrystals during the condensation process. This behaviour is the opposite of that previously reported for the addition of salt to NCC, which is believed to reduce the helical pitch by masking electrostatic repulsion<sup>27</sup>.

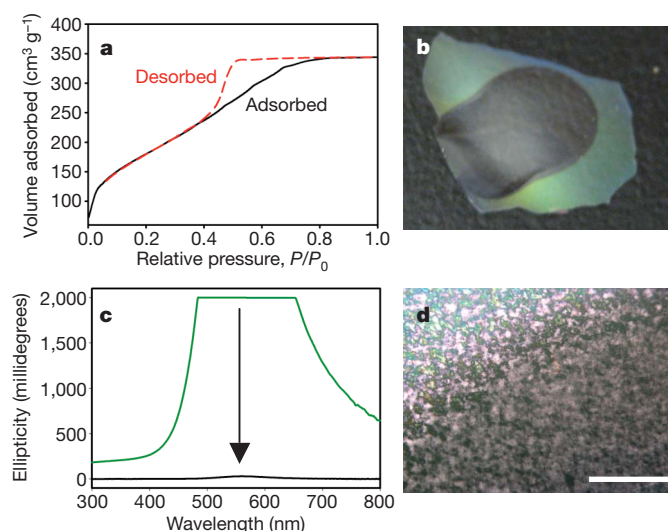
Removal of the cellulose template was accomplished by calcination of the composite films at 540 °C under air (Supplementary Figs 3–5) and resulted in free-standing mesoporous silica films (Fig. 2c). The calcined films show strong birefringence by POM and a texture that is very similar to that observed for the composite films (Fig. 1d). The

reflection peaks of the four composite films shown in Fig. 2a are blue-shifted by approximately 300 nm after pyrolysis to give iridescent silica films that reflect light at different wavelengths across the entire visible spectrum (Fig. 2b–e). Overall, by starting with different composite films, mesoporous silica films with reflectance peaks at wavelengths ranging from ~300 to 1,300 nm were obtained. The blueshift that occurs after calcination is greater than that expected from the calculated decrease in  $n_{\text{avg}}$  caused by the removal of cellulose alone, and is probably because of (1) contraction of the films, resulting in a shorter helical pitch, and (2) the decrease in  $n_{\text{avg}}$  (see Supplementary Discussion and Supplementary Table 2 for quantitative details). Circular dichroism experiments showed a strong positive signal with intensity greater than 2,000 millidegrees for all of the coloured films (Supplementary Fig. 8). This demonstrates that the observed colours arise from the selective reflection of left-handed polarized light and confirms that the left-handed chiral nematic structure from NCC is preserved in the mesoporous silica films.

The silica films are mesoporous, as determined by nitrogen adsorption studies (Fig. 3a). Type IV adsorption isotherms with type H2 hysteresis loops are observed in all of the calcined samples, with Brunauer–Emmett–Teller (BET) surface areas in the range ~800–300 m<sup>2</sup> g<sup>−1</sup> and pore volumes of ~0.60–0.25 cm<sup>3</sup> g<sup>−1</sup>, depending on the NCC/silica ratio (Supplementary Table 1). The BJH (Barrett, Joyner and Halenda) pore size distributions give a peak pore diameter of 3.5–4 nm and show very little pore volume past 8 nm (Supplementary Fig. 7). This corresponds well to a previously reported diameter of 5 nm for wood-based NCC (also see Supplementary Fig. 1)<sup>21</sup>, thus showing that individual nanocrystals, as opposed to bundles, were successfully replicated in the pore structure. Transmission electron microscopy (TEM) was also used to study the pores of the chiral nematic mesoporous silica materials (Supplementary Fig. 2). Long cylindrical pores with an organization that is consistent with a chiral nematic structure are seen by TEM. The pores appear to have fairly uniform diameters of ~4–5 nm, which is in good agreement with the nitrogen adsorption data.



**Figure 2 | Optical characterization of NCC/silica composite films and the corresponding mesoporous silica films.** **a**, Transmission spectra of four NCC/silica composite films with reflectance peaks in the near-infrared part of the spectrum. The proportion of TMOS:NCC was increased from samples S1 to S4, resulting in a redshift in the reflectance peaks of the films. **b**, Transmission spectra of the mesoporous silica films obtained from the calcination of composite films S1 to S4. The reflectance peaks were all blueshifted by approximately 300 nm, resulting in films that reflect light across the entire visible spectrum. **c**, Photograph showing the different colours of mesoporous silica films S1 to S4. The colours in these silica films arise only from the chiral nematic pore structure present in the materials. The dime is included for scale (diameter, 18 mm). **d**, Photograph of a yellow mesoporous silica film (S3) taken at normal incidence. **e**, Photograph of the same film taken at oblique incidence appears blue owing to the  $\sin\theta$  dependence of the reflected wavelength.



**Figure 3 | Nitrogen and water absorption in a chiral mesoporous silica film.** **a**, Type IV adsorption isotherm measured for a chiral nematic mesoporous silica film (N<sub>2</sub>/77 K). **b**, Photograph of a green mesoporous silica film (S2) after the addition of a drop of water, which causes the wet part of the film to become completely transparent. **c**, The circular dichroism spectra of a green mesoporous silica film before (green curve) and after (black curve) infiltration with water, which results in a reversible decrease of the circular dichroism signal to 30 millidegrees. **d**, POM image of a green mesoporous silica film after the addition of a drop of water showing almost complete loss of birefringence in the wet part of the film compared to the region that is still dry (top left corner) (scale bar, 300 μm).



SEM provided further confirmation of the replication of chiral nematic organization in the pure silica films (Fig. 4). The top surface of the films appear very smooth, but perpendicular to the surface we observe a layered structure with defects that arise from changes in direction of the helical axis of the chiral nematic phase (Fig. 4a). This is expected because no macroscopic alignment was performed on the samples, and is consistent with the domains observed by POM. At higher magnification we can see that the repeating distance is of the order of several hundred nanometres, which is in agreement with the reflection of visible light (Fig. 4b and c). At very high magnification we can resolve a twisting rod-like morphology (Fig. 4d). Throughout the sample, this twisting appears to occur in a counter-clockwise direction when moving away from the viewer, consistent with a left-handed helical organization. In some locations defects can be seen that correspond to a condensed version of those observed by POM in the liquid crystal phase (Fig. 4e and f). This defect structure is also very similar to that recently observed for chiral iridescent beetle exoskeletons<sup>19</sup>. In general these SEM images look very similar to those obtained for the composite films and pure NCC films (Supplementary Figs 1b and 6). We therefore have direct evidence that the chiral nematic organization of NCC has been faithfully replicated in the mesoporous silica films.

To demonstrate a unique property of the chiral nematic mesoporous silicas, we examined their absorption of isotropic liquids. These films rapidly absorb water and become completely transparent and colourless, which can be observed visually (Fig. 3b). A similar effect has also been reported for engineered helical inorganic nanostructures, and may be attributed to approximate refractive index matching between

the isotropic liquid in the pores and the silica walls<sup>28</sup>. As a control, no change is apparent when water (or other common solvents) is added to an NCC/silica composite film before calcination. The birefringence of the mesoporous films is also drastically reduced when a solvent is absorbed (Fig. 3d). These changes are reversible and the films fully regain their iridescence and birefringence upon drying. The ability to switch between iridescent and colourless films combined with the wide tunability of helical pitch suggests that these materials could find use in smart window applications.

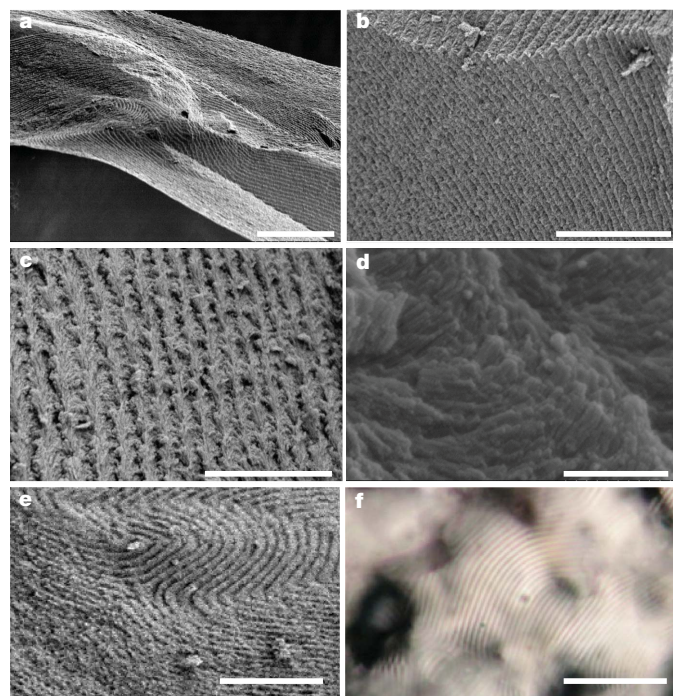
Because the refractive indices of water ( $n = 1.33$ ) and  $\text{SiO}_2$  ( $n = 1.46$ ) are not perfectly matched, a small residual reflectance peak is expected after water infiltration; it is too small, however, to be detected by the naked eye or in the transmission spectra of the water-soaked mesoporous films. The chiral origin of the reflectance peak allows us to probe optical changes using circular dichroism, in which the signal is reduced by two orders of magnitude to 30 millidegrees after infiltration with water (Fig. 3c). The circular dichroism peak is also redshifted compared to the reflectance peak for the dry film, owing to the increase in  $n_{\text{avg}}$ . Infiltration with isopropanol ( $n = 1.38$ ) gives a circular dichroism signal that is further redshifted and considerably less intense than that observed for water, based on a fairly small difference in refractive index (Supplementary Fig. 9). Absorption of dimethyl sulphoxide into the films ( $n = 1.48$ ) completely eliminates the circular dichroism signal, owing to almost perfect refractive index matching with  $\text{SiO}_2$ . These results suggest that there are opportunities to employ these materials in optical sensing devices where small changes in the refractive index within the pores would result in changes in both the intensity and position of the circular dichroism peak. This takes advantage of the unique combination of chirality, optical properties, and mesoporosity in these materials and the sensitivity of circular dichroism.

We have shown, for the first time, that free-standing mesoporous silica films with long-range chiral nematic ordering may be obtained using a template-based approach. The chiral nematic organization and high surface area of nanocrystalline cellulose is accurately replicated in the inorganic solid. The helical structure of the mesoporous films results in chiral reflectance that can be tuned across the entire visible spectrum and into the near-infrared. Along with the potential applications described above, we believe that these materials can be used as hard templates to synthesize a variety of new materials with chiral nematic structures that have hitherto been inaccessible. In addition, because the mesoporous silica materials are imprints of chiral cellulose particles at multiple levels, and cellulose is a leading material for chiral separation, these materials may be useful for separating enantiomers. Finally, the wide availability and renewable nature of cellulose, combined with the simplicity of our approach, suggest that bulk quantities of chiral nematic mesoporous silica could be economically produced in a sustainable fashion.

## METHODS SUMMARY

**Preparation of chiral nematic mesoporous silica.** In a typical procedure, chiral nematic mesoporous silica was prepared by first sonicating 10 ml of a 3% aqueous NCC suspension (pH 2.4) for 10 min (see Methods for details of NCC preparation). TEOS (0.60 ml, 2.7 mmol) was added to the NCC suspension and the mixture was stirred at 60 °C until a homogeneous mixture was obtained (typically about 3 h). This solution was cooled to room temperature, then allowed to dry on a polystyrene Petri dish. (Alternatively, an equivalent molar amount of TMOS was used, in which case the mixture was stirred at room temperature for 1 h before evaporation.) After slow evaporation at room temperature, free-standing films of the NCC/silica composite materials were obtained (490 mg).

For pyrolysis of the cellulose, under flowing air the composite film (300 mg) was heated at a rate of 120 °C h<sup>-1</sup> to 100 °C, held at that temperature for 2 h, then heated to 540 °C at 120 °C h<sup>-1</sup> and held at that temperature for 6 h. After slowly cooling to room temperature, 100 mg of free-standing films were recovered. Additional samples were prepared by varying the ratio of TEOS/TMOS:NCC used in the synthesis. In general, materials prepared with TMOS showed better film quality.



**Figure 4 | SEM images of chiral nematic mesoporous silica films and comparison of fingerprint textures in the solid state and liquid crystal phase.** **a**, Top view of a cracked film shows the relatively smooth top surface and a layered structure looking down the edge (scale bar, 10  $\mu\text{m}$ ). **b**, Side view of a cracked film shows the stacked layers that result from the helical pitch of the chiral nematic phase (scale bar, 3  $\mu\text{m}$ ). **c**, Higher magnification reveals the helical pitch distance to be of the order of several hundred nanometres (scale bar, 2  $\mu\text{m}$ ). **d**, Very high magnification shows a rod-like morphology with the rods twisting in a left-handed orientation (scale bar, 200 nm). **e**, Fingerprint defect in a solid mesoporous silica film (scale bar, 10  $\mu\text{m}$ ). **f**, Fingerprint defect observed by POM in the liquid crystal phase of an NCC/TEOS mixture (scale bar, 30  $\mu\text{m}$ ).



**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 27 April; accepted 28 September 2010.**

- Johnson, L. N. Asymmetry at the molecular level in biology. *Eur. Rev.* **13**, 77–95 (2005).
- Gabashvili, A., Medina, D. D., Gedanken, A. & Mastai, Y. Templating mesoporous silica with chiral block copolymers and its application for enantioselective separation. *J. Phys. Chem. B* **111**, 11105–11110 (2007).
- Johnson, B. F. G. *et al.* Superior performance of a chiral catalyst confined within mesoporous silica. *Chem. Commun.* 1167–1168 (1999).
- Fireman-Shoresh, S., Popov, I., Avnir, D. & Marx, S. Enantioselective, chirally templated sol-gel thin films. *J. Am. Chem. Soc.* **127**, 2650–2655 (2005).
- Hodgkinson, I. & Wu, Q. H. Inorganic chiral optical materials. *Adv. Mater.* **13**, 889–897 (2001).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Yang, P., Zhao, D., Margolese, D. I., Chmelka, B. F. & Stucky, G. D. Generalized syntheses of large-pore mesoporous metal oxides with semicrystalline frameworks. *Nature* **396**, 152–155 (1998).
- Yang, H., Coombs, N., Sokolov, I. & Ozin, G. A. Free-standing and oriented mesoporous silica films grown at the air–water interface. *Nature* **381**, 589–592 (1996).
- Armatas, G. S. & Kanatzidis, M. G. Hexagonal mesoporous germanium. *Science* **313**, 817–820 (2006).
- Inagaki, S., Guan, S., Ohsuna, T. & Terasaki, O. An ordered mesoporous organosilica hybrid material with a crystal-like wall structure. *Nature* **416**, 304–307 (2002).
- Sun, D. *et al.* Hexagonal nanoporous germanium through surfactant-driven self-assembly of Zintl clusters. *Nature* **441**, 1126–1130 (2006).
- Attard, G. S., Glyde, J. C. & Göltner, C. G. Liquid-crystalline phases as templates for the synthesis of mesoporous silica. *Nature* **378**, 366–368 (1995).
- Che, S. *et al.* Synthesis and characterization of chiral mesoporous silica. *Nature* **429**, 281–284 (2004).
- Qiu, H., Inoue, Y. & Che, S. Supramolecular chiral transcription and recognition by mesoporous silica prepared by chiral imprinting of a helical micelle. *Angew. Chem. Int. Ed.* **48**, 3069–3072 (2009).
- Broer, D. J., Lub, J. & Mol, G. N. Wide-band reflective polarizers from cholesteric polymer networks with a pitch gradient. *Nature* **378**, 467–469 (1995).
- Yang, D.-K., West, J. L., Chien, L.-C. & Doane, J. W. Control of reflectivity and bistability in displays using cholesteric liquid crystals. *J. Appl. Phys.* **76**, 1331–1333 (1994).
- Kopp, V. I., Fan, B., Vithana, H. K. M. & Genack, A. Z. Low-threshold lasing at the edge of a photonic stop band in cholesteric liquid crystals. *Opt. Lett.* **23**, 1707–1709 (1998).
- Akagi, K. *et al.* Helical polyacetylene synthesized with a chiral nematic reaction field. *Science* **282**, 1683–1686 (1998).
- Sharma, V., Cme, M., Park, J. O. & Srinivasarao, M. Structural origin of circularly polarized iridescence in jeweled beetles. *Science* **325**, 449–451 (2009).
- Mukherjee, S. M. & Woods, H. J. X-ray and electron microscope studies of the degradation of cellulose by sulphuric acid. *Biochim. Biophys. Acta* **10**, 499–511 (1953).
- Revol, J.-F., Bradford, H., Giasson, J., Marchessault, R. H. & Gray, D. G. Helicoidal self-ordering of cellulose microfibrils in aqueous suspension. *Int. J. Biol. Macromol.* **14**, 170–172 (1992).
- Revol, J.-F., Godbout, L. & Gray, D. G. Solid self-assembled films of cellulose with chiral nematic order and optically variable properties. *J. Pulp Pap. Sci.* **24**, 146–149 (1998).
- Dujardin, E., Blaseby, M. & Mann, S. Synthesis of mesoporous silica by sol-gel mineralisation of cellulose nanorod nematic suspensions. *J. Mater. Chem.* **13**, 696–699 (2003).
- Thomas, A. & Antonietti, M. Silica nanocasting of simple cellulose derivatives: towards chiral pore systems with long-range order and chiral optical coatings. *Adv. Funct. Mater.* **13**, 763–766 (2003).
- Edgar, C. D. & Gray, D. G. Induced circular dichroism of chiral nematic cellulose films. *Cellulose* **8**, 5–12 (2001).
- De Vries, H. L. Rotatory power and other optical properties of certain liquid crystals. *Acta Crystallogr.* **4**, 219–226 (1951).
- Dong, X. M., Kimura, T., Revol, J.-F. & Gray, D. G. Effects of ionic strength on the isotropic-chiral nematic phase transition of suspensions of cellulose crystallites. *Langmuir* **12**, 2076–2082 (1996).
- Robbie, K., Broer, D. J. & Brett, M. J. Chiral nematic order in liquid crystals imposed by an engineered inorganic nanostructure. *Nature* **399**, 764–766 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and FPLInnovations. K.E.S. is grateful to UBC for a graduate fellowship.

**Author Contributions** K.E.S. conducted all of the synthesis and most of the characterization. H.Q. assisted with characterization of the materials. W.Y.H. supplied the NCC and contributed valuable expertise on NCC. M.J.M. initiated and guided this work.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.J.M. ([mmacclach@chem.ubc.ca](mailto:mmacclach@chem.ubc.ca)).

## METHODS

**Preparation of nanocrystalline cellulose.** For the preparation of nanocrystalline cellulose (NCC) fully-bleached, commercial kraft softwood pulp was first milled to pass through a 0.5-mm screen in a Wiley mill to ensure particle size uniformity and to increase surface area. The milled pulp was hydrolysed in sulphuric acid (8.75 ml of a sulphuric acid solution per gram of pulp) at a concentration of 64 wt% and a temperature of 45 °C with vigorous stirring for 25 min. The cellulose suspension was then diluted with cold de-ionized water (about ten times the volume of the acid solution used) to stop the hydrolysis, and allowed to settle overnight. The clear top layer was decanted and the remaining cloudy layer was centrifuged. The supernatant was decanted and the resulting thick white suspension was washed three times with de-ionized water to remove all soluble cellulose materials. The thick white suspension obtained after the last centrifugation step was placed inside dialysis membrane tubes (12,000–14,000 molecular weight cut-off) and dialysed against slow running de-ionized water, for 1 to 4 days. The membrane tubes containing the extracted cellulose materials were placed periodically in de-ionized H<sub>2</sub>O, and the procedure was continued until the pH of the water became constant for a period of one hour. The suspension from the membrane tubes was dispersed by subjecting it to ultrasound treatment in a Fisher Sonic Dismembrator (Fisher Scientific) for 10 min at 60% power and then diluted to the desired concentration.

**Preparation of chiral nematic mesoporous silica.** In a typical procedure, chiral nematic mesoporous silica was prepared by first sonicating 10 ml of a 3% aqueous NCC suspension (pH 2.4) for 10 min. TEOS (0.60 ml, 2.7 mmol) was added to the NCC suspension and the mixture was stirred at 60 °C until a homogeneous mixture was obtained (typically about 3 h). This solution was cooled to room temperature, then allowed to dry on a polystyrene Petri dish. (Alternatively, an equivalent molar amount of TMOS was used, in which case the mixture was stirred at room temperature for 1 h before evaporation.) After slow evaporation at room temperature, free-standing films of the NCC/silica composite materials were obtained (490 mg).

For pyrolysis of the cellulose, under flowing air the composite film (300 mg) was heated at a rate of 120 °C h<sup>-1</sup> to 100 °C, held at that temperature for 2 h, then

heated to 540 °C at 120 °C h<sup>-1</sup> and held at that temperature for 6 h. After slowly cooling to room temperature, 100 mg of free-standing films were recovered. Additional samples were prepared by varying the ratio of TEOS/TMOS:NCC used in the synthesis. In general, materials prepared with TMOS showed better film quality.

**STEM of nanocrystalline cellulose.** STEM images were obtained on a FEI Quanta 400F environmental SEM. To image the individual NCC crystallites, a 1 µl drop of approximately 0.005 wt% solution was deposited on a TEM grid immediately after sonication (for at least 20 min) and dried under ambient conditions. Pre-made 400 mesh copper Formvar and carbon-coated grids from Pacific Grid Tech were used.

**Characterization of composite films and chiral nematic mesoporous silica.**

Ultraviolet-visible/near-infrared spectroscopy was conducted on a Cary 5000 UV-Vis/NIR spectrophotometer. Transmission spectra were collected by mounting free-standing films so that the surfaces of the films were perpendicular to the beam path. The maximum transmittance was set to 100% in a region away from the reflectance peak. Polarized optical microscopy was performed on an Olympus BX41 microscope. All images were taken with the polarizers in a perpendicular (crossed) arrangement. Thermogravimetric analysis was performed on a PerkinElmer Pyris 6 thermogravimetric analyser. Infrared spectra were obtained with a Nicolet 6700 FT-IR equipped with a Smart Orbit diamond attenuated total reflectance (ATR) attachment. Powder X-ray diffraction spectra were collected using a D8 advance X-ray diffractometer. TEM images were collected on a Hitachi H7600 electron microscope. Samples were prepared by first grinding the films into a fine powder and then dropcasting them onto a copper TEM grid. SEM images were collected on a Hitachi S4700 electron microscope. Samples were prepared by breaking films into small pieces and attaching them to aluminium stubs using double-sided adhesive tape. The samples were then sputter-coated with either gold or gold-palladium. Gas adsorption studies were performed using a Micromeritics Accelerated Surface Area & Porosity (ASAP) 2000 system. Circular dichroism experiments were performed using a JASCO J-710 spectropolarimeter. Spectra were collected by mounting free-standing films so that the surfaces of the films were perpendicular to the beam path.

# Intrusion triggering of the 2010 Eyjafjallajökull explosive eruption

Freysteinn Sigmundsson<sup>1</sup>, Sigrún Hreinsdóttir<sup>1</sup>, Andrew Hooper<sup>2</sup>, Thóra Árnadóttir<sup>1</sup>, Rikke Pedersen<sup>1</sup>, Matthew J. Roberts<sup>3</sup>, Niels Óskarsson<sup>1</sup>, Amandine Auriac<sup>1</sup>, Judicael Decrie<sup>1</sup>, Páll Einarsson<sup>1</sup>, Halldór Geirsson<sup>3</sup>, Martin Hensch<sup>1</sup>, Benedikt G. Ófeigsson<sup>1</sup>, Erik Sturkell<sup>4</sup>, Hjörleifur Sveinbjörnsson<sup>3</sup> & Kurt L. Feigl<sup>5</sup>

Gradual inflation of magma chambers often precedes eruptions at highly active volcanoes. During such eruptions, rapid deflation occurs as magma flows out and pressure is reduced<sup>1–3</sup>. Less is known about the deformation style at moderately active volcanoes, such as Eyjafjallajökull, Iceland, where an explosive summit eruption of trachyandesite beginning on 14 April 2010 caused exceptional disruption to air traffic, closing airspace over much of Europe for days. This eruption was preceded by an effusive flank eruption of basalt from 20 March to 12 April 2010. The 2010 eruptions are the culmination of 18 years of intermittent volcanic unrest<sup>4–9</sup>. Here we show that deformation associated with the eruptions was unusual because it did not relate to pressure changes within a single magma chamber. Deformation was rapid before the first eruption (>5 mm per day after 4 March), but negligible during it. Lack of distinct co-eruptive deflation indicates that the net volume of magma drained from shallow depth during this eruption was small; rather, magma flowed from considerable depth. Before the eruption, a ~0.05 km<sup>3</sup> magmatic intrusion grew over a period of three months, in a temporally and spatially complex manner, as revealed by GPS (Global Positioning System) geodetic measurements and interferometric analysis of satellite radar images. The second eruption occurred within the ice-capped caldera of the volcano, with explosivity amplified by magma–ice interaction. Gradual contraction of a source, distinct from the pre-eruptive inflation sources, is evident from geodetic data. Eyjafjallajökull's behaviour can be attributed to its off-rift setting with a 'cold' subsurface structure and limited magma at shallow depth, as may be typical for moderately active volcanoes. Clear signs of volcanic unrest signals over years to weeks may indicate reawakening of such volcanoes, whereas immediate short-term eruption precursors may be subtle and difficult to detect.

Volcanic processes leading to eruptions can be investigated by monitoring a variety of phenomena, including increased earthquake activity and uplift of volcanoes due to magma accumulation, as well as variations in heat and gas emission. Many well-documented cases of pre-eruptive changes come from some of the more active volcanoes on Earth, whereas such observations at long-dormant volcanoes are limited. Because there is little magmatic heat input, the edifice and internal structure of a long-dormant volcano may be colder than at more active ones. Are magma movements in the roots of such 'cold' volcanoes different from those at more active volcanoes? Detailed deformation measurements over two decades at the Eyjafjallajökull volcano in south Iceland help to answer this question, demonstrating how transfer of magma inside the volcano led to eruptive activity after almost two centuries of quiescence.

Eyjafjallajökull volcano is situated in a propagating rift outside the main zone of plate spreading in Iceland, at the southern termination of the eastern rift zone (Fig. 1). The area is characterized by high-rising

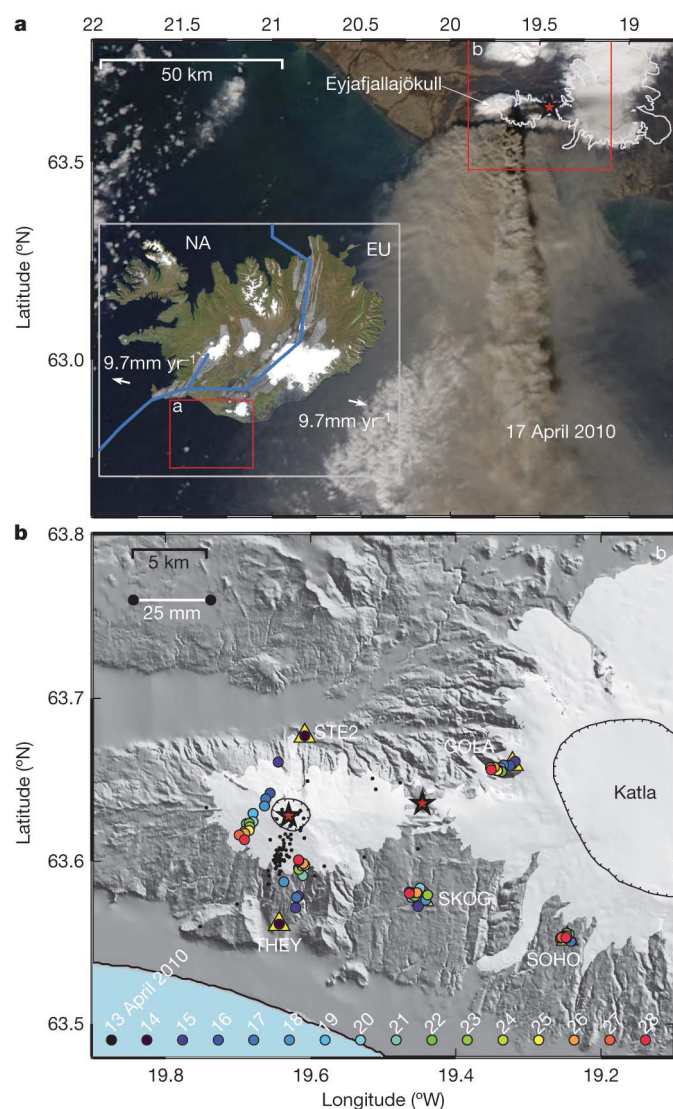
central volcanoes but lacks rift structures that result from plate spreading. Magma generation differs from that in the rift zones<sup>10</sup>. Previous eruptions of Eyjafjallajökull include a radial fissure eruption around AD 920, a small summit eruption in AD 1612 or 1613, and another summit eruption in AD 1821–23. A short phreato-magmatic phase in December 1821 was followed by a year-long period of intermittent magmatic/phreato-magmatic activity and flooding (ref. 11, and G. Larsen, personal communication). An explosive eruption that began on 14 April 2010 was the culmination of a long series of intermittent magmatic events observed over 18 years. In 1992, earthquake activity increased at the volcano following over 20 years of quiescence since measurements began. Extensive intrusions formed beneath the volcano in 1994 and 1999 (refs 4–8). Deformation associated with the events was mapped by interferometric satellite radar (InSAR) observations, GPS geodetic measurements, and optical tilt levelling. Studies of these data reveal sill intrusions at 4.5–6.5 km depth as the most likely source of deformation. The 1994 and 1999 intrusions had inferred volumes of  $\sim(10\text{--}17) \times 10^6 \text{ m}^3$  and  $\sim(21\text{--}31) \times 10^6 \text{ m}^3$ , respectively<sup>4–6</sup>. Between mid-2000 and 2009, earthquakes occurred intermittently at rates of 1–4 events per month, while deformation remained negligible.

In mid-2009, seismicity and deformation picked up again for a period of a few weeks, with 10–12 mm of southward displacement at GPS station THEY, located on the south side of the volcano. At the beginning of January 2010, deformation was again detected, and the level of seismicity increased to several earthquakes per day<sup>12</sup> (Supplementary Fig. 1). These changes marked the onset of magma flow into the roots of the volcano, culminating in late evening 20 March 2010 with the opening of a short effusive fissure on the volcano's flank, which erupted basaltic lava with a SiO<sub>2</sub> content of 48% (see Methods). Remarkably, deformation almost ceased when the vents opened and the volcano remained at an inflated stage, without significant subsidence until 9 April. Then subsidence occurred at GPS station STE2, three days before the end of the flank eruption on 12 April. This eruption produced lava at an average rate of  $\sim 13 \text{ m}^3 \text{ s}^{-1}$ , or  $25 \times 10^6 \text{ m}^3$  in total<sup>13</sup>. Following a 1–2 day hiatus in eruptive activity, during which STE2 showed renewed inflation, an explosive eruption began on 14 April when a new set of vents formed at the ice-capped summit of the volcano. Sustained, highly variable activity continued until 22 May with an average magma eruption rate of  $30\text{--}60 \text{ m}^3 \text{ s}^{-1}$ ; a pronounced peak in activity occurred during the initial four days<sup>13</sup>. Initially analysed samples have SiO<sub>2</sub> content of about 58% and are classified as trachyandesite. Interaction of magma and ice initially augmented explosive activity, generating fine-grained tephra that rose to heights of 6–9 km.

Deformation of the Eyjafjallajökull volcano associated with the 2010 eruptions has been measured by two complementary geodetic techniques. InSAR reveals the spatial extent of deformation and its cumulative amount between image acquisitions. Temporal resolution of deformation is provided by analyses of data from continuously recording

<sup>1</sup>Nordic Volcanological Center, Institute of Earth Sciences, University of Iceland, Askja, Sturlugata 7, Reykjavik IS-101, Iceland. <sup>2</sup>Delft Institute of Earth Observation and Space Systems, Delft University of Technology, Kluyverweg 1, 2629 HS, Delft, The Netherlands. <sup>3</sup>Icelandic Meteorological Office, Bustadavegur 9, IS-150 Reykjavik, Iceland. <sup>4</sup>Department of Earth Sciences, University of Gothenburg, Box 460, SE-405 30 Gothenburg, Sweden. <sup>5</sup>Department of Geoscience, University of Wisconsin-Madison, 1215 W. Dayton St., Madison, Wisconsin 53706, USA.





**Figure 1 | Iceland and the location of Eyjafjallajökull volcano.** **a**, Main panel, satellite image (MODIS) of the eruption plume emanating from Eyjafjallajökull main crater on 17 April. Outline of ice caps shown with grey line. Red star, location of the preceding flank eruption. Red box, location of **b**. Inset, satellite image of Iceland overlain by fissure swarms in rift zones (shaded grey) and showing schematically in blue the main axes of the plate boundary between the North American (NA) and Eurasian (EU) plates. Half-spreading rate is  $9.7 \text{ mm yr}^{-1}$ . Red box, location of **a**. **b**, Locations of GPS stations (triangles) operating before the 2010 eruptive activity around Eyjafjallajökull. Daily station horizontal displacements 14–28 April are shown in colour (dots at bottom show colour-coding of dates). Stations closest to the volcano contract towards it. Earthquake epicentres 14–29 April are indicated by black dots. The summit caldera of Eyjafjallajökull volcano is indicated, as well as the larger caldera of the neighbouring Katla volcano. Red stars show the two eruptive sites. Image in **a** is courtesy of NASA/GSFC, MODIS Rapid Response (<http://rapidfire.sci.gsfc.nasa.gov/>).

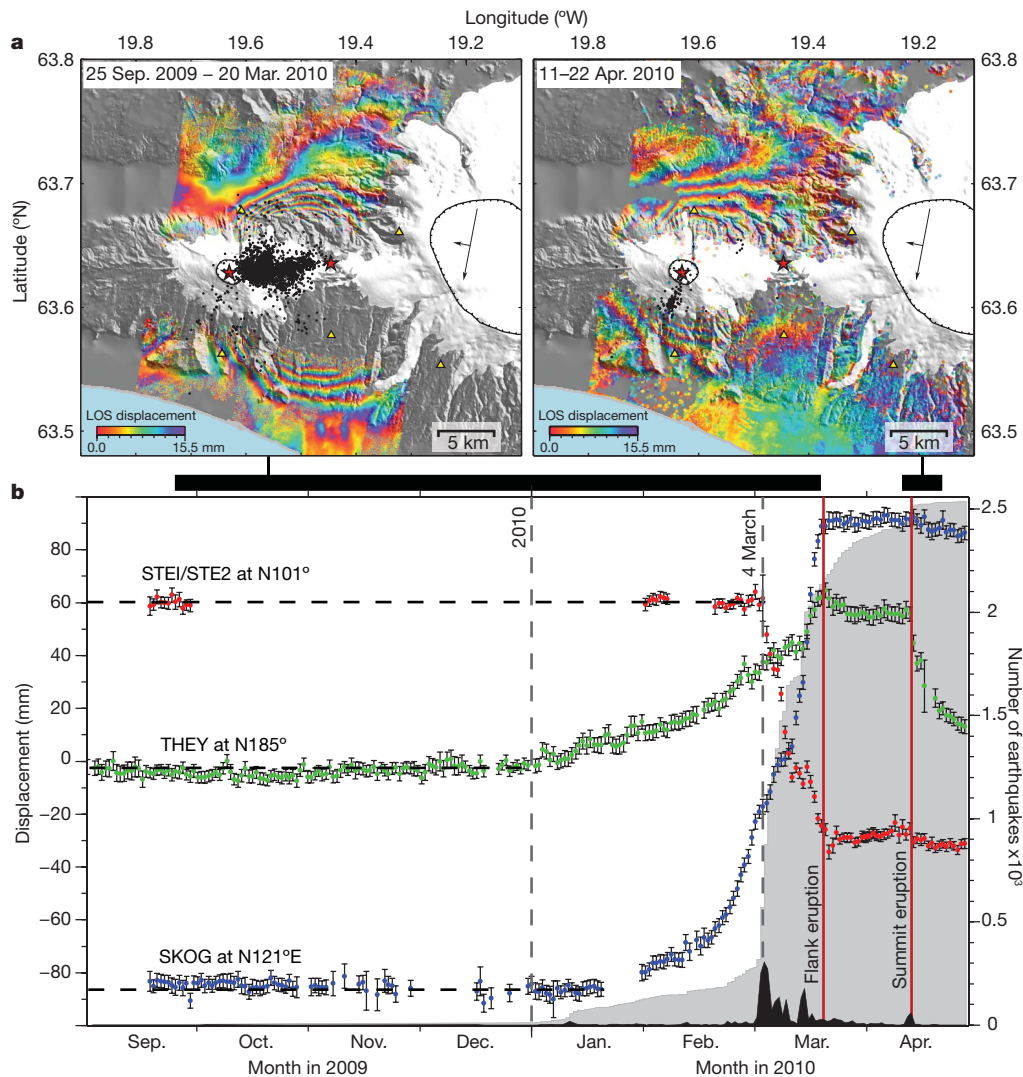
GPS geodetic receivers around the volcano that give three-dimensional displacements as a function of time. We asked the German Space Agency (DLR) to program the TerraSAR-X satellite<sup>14</sup> to acquire images over Eyjafjallajökull, beginning in July 2009. On 20 March 2010, images were acquired in both ascending and descending satellite tracks, providing interferograms spanning almost the entire pre-eruptive inflation interval until a few hours before the first eruption. InSAR interferograms spanning the flank eruption indicate no detectable deformation, whereas those spanning the initial part of the second eruption show deflation (Fig. 2 and Supplementary Fig. 3). The GPS data reveal a complex pattern. Station THEY began to move southward in January, indicating

inflation. Station SKOG began to move southeastwards a few weeks later. A final phase of pre-eruptive deformation began on 4 March when station STE2 started moving westward. In combination, the geodetic data reveal three stages of deformation: (1) a pre-eruptive stage of inflation due to a complicated time-evolving magma intrusion that produced variable and high rates of deformation, in particular after 4 March; (2) from 20 March to 9 April, a co-eruptive stage characterized by a pause in deformation (negligible rates); and (3) a co-eruptive deformation stage associated with the April–May summit eruption, indicating gradual deflation of a source distinct from the pre-eruptive inflation source.

Modelling indicates a pre-eruptive intrusion of complicated geometry over an extended depth range. The combined GPS and InSAR data set for this period is poorly fitted by models involving a single spherical or tabular intrusion (dyke or sill). Inversion has been conducted to find sources embedded within a homogeneous elastic half-space capable of recreating the observed deformation. Two different approaches were used, based on: (1) sources of simple geometries<sup>15,16</sup> and (2) sources of irregular shape and variable opening, imposing a hydrostatic overpressure boundary condition (Methods Summary). A Markov chain Monte Carlo method was used to estimate the probability distribution of the model parameters<sup>17</sup>. Both methods give the maximum likelihood range for the model parameters (95% confidence intervals). Observations from December to the end of February can be explained by a single horizontal sill inflating at a depth of 4.0–5.9 km under the southeastern flank of the volcano, with a volume increase of  $(8\text{--}18) \times 10^6 \text{ m}^3$ . The March pre-eruptive deformation is explained by a second sill at about the same depth under the northeastern flank of the volcano, together with a southeast-tilted dyke reaching from 3.2–6.1 km depth to within a few hundred metres, or less, of the surface (Fig. 3). The sills link to a seismically inferred ‘upflow zone’ of magma within the volcano<sup>12,18,19</sup>, and the dyke model is consistent with its origin at the depth of the sills. The modelling thus suggests one interconnected intrusion in the pre-eruptive stage with magma flow at an average rate of  $2\text{--}3 \text{ m}^3 \text{ s}^{-1}$  before March, followed by an average flow rate of  $30\text{--}40 \text{ m}^3 \text{ s}^{-1}$ . The total inferred volume increase is  $(49\text{--}71) \times 10^6 \text{ m}^3$ . A deflating source associated with the 14 April summit eruption is simpler and better constrained than the pre-eruptive intrusion. The data can be fitted by a horizontal deflating sill at a depth of 4.0–4.7 km, with a volume decrease of  $(13\text{--}15) \times 10^6 \text{ m}^3$  until 22 April. This source is spatially offset from, and different to, the pre-eruptive complex.

Our observations have revealed the growth of an intrusive complex in the roots of Eyjafjallajökull volcano during the three months before eruption onset. After initial horizontal growth, it shows both horizontal and sub-vertical growth in the three weeks before the first eruption. This behaviour can be attributed to subsurface variations in crustal stress and strength originating from complicated volcano foundations formed in a propagating rift. A low-density layer may capture magma, allowing pressure to build before an intrusion can ascend towards higher levels<sup>20</sup>. The intrusive complex was presumably formed by basalt, as erupted on the volcano flank from 20 March to 12 April; its growth halted at the onset of this eruption. Deformation associated with the eruption onset was minor, as the dyke had already approached the surface. Furthermore, our inferred average inflow rate to the intrusion in March is about three times higher than the average eruption rate during the flank eruption. Isolated eruptive vents opening on long-dormant volcanoes may represent magma leaking upwards from extensive pre-eruptive intrusions formed at depth. The deflation source activated during the summit eruption of trachyandesite is distinct from, and adjacent to, all documented sources of inflation in the volcano roots.

The basaltic magma that recharged the volcano appears to have triggered the summit eruption, although the exact mode of triggering is uncertain. Scenarios include stress triggering or propagation of basalt into more evolved magma. The trachyandesite includes crystals that could be remnants of a minor recent intrusion of basalt (see Methods). The trachyandesite may have formed over many years



**Figure 2 | InSAR, GPS and seismic data at Eyjafjallajökull.** **a**, TerraSAR-X interferograms from descending satellite orbits, spanning the pre-eruptive intrusive period (left, time period 25 September 2009 to 20 March 2010 at 7:49 GMT) and the initial days of the explosive eruption (right, time period 11 to 22 April 2010). Black orthogonal arrows show the satellite flight path and look direction. One colour fringe corresponds to line-of-sight (LOS) change of 15.5 mm (positive for increasing range, that is, motion of the ground away from the satellite). Black dots show earthquake epicentres for the corresponding period. Background is shaded topography. Thick lines below indicate the time span of the interferograms. Red stars and triangles same as in Fig. 1. **b**, Selected displacement components from 1 September 2009 to 28 April 2010 at GPS

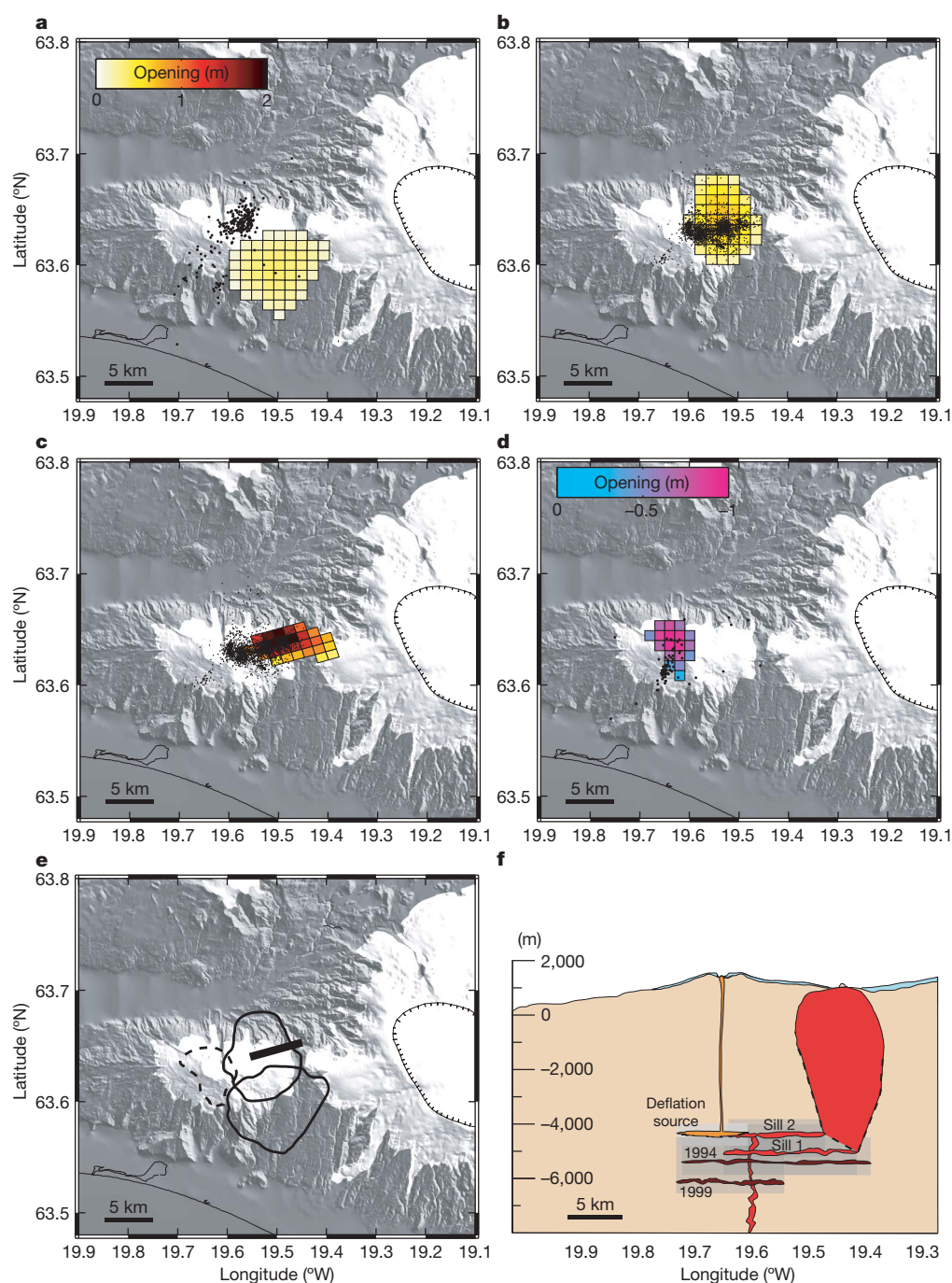
stations THEY, SKOG and STEI/STE2 (see Fig. 1 for locations). The curves are offset, to plot on the same panel. Error bars,  $1\sigma$  confidence intervals. Displacement in direction N185°E is shown for station THEY and direction N121°E for SKOG, both away from the summit of Eyjafjallajökull. Displacement in direction N101°E is shown for STEI/STE2, perpendicular to the direction to the summit. (STEI and STE2 are two stations co-located 3.9 m apart and recording the same motion. The displacement labelled STEI/STE2 shows results from STEI up until 6 February; after this date results are from STE2.) Full time series are shown in Supplementary Fig. 2. Gray shading shows the cumulative number of earthquakes and black shading the corresponding daily rate; for full details of seismicity see Methods.

before the triggering intrusion, by partial melting of hydrated crust around the margins of multiple basaltic intrusions, within a domain at the melting point of the trachyandesite. Alternatively, mixing of larger portion of basalt with more evolved magma may have occurred<sup>21</sup>. The geodetically inferred sill-shaped source of deflation during the eruption suggests, however, that magma was drained from a widespread domain under the summit area, eventually limiting the flow rate and contributing to the long duration of the summit eruption. Considering uncertainties in the modelling, we cannot clearly discriminate between a model with a stack of vertically separated sills and a model consisting of all the intrusions occurring at the same depth. A layer of partial melt or magma within the volcano would contribute to capturing of basalt intrusions from depth, because of its low strength and density. The geodetic data show that the intrusion formed before the initial eruption did not deflate during the second one. This argues against all the

deformation sources relating to a single magma chamber. The short eruptive pause following the closure of the flank eruption's feeder channel suggests critical pressure conditions in the system. Continued inflow from depth caused eventual pressure build-up in the intrusive complex, providing the trigger needed to initiate the explosive eruption.

Our observations have implications for interpretation of other cases of bimodal volcanism, such as the 1996 eruption of Karymsky volcano when andesite erupted from its main vent and basalt from its flank<sup>22</sup>, and for studies of eroded exposures of complicated volcano interiors that reveal interconnected network of sills and dykes<sup>23</sup>. Intrusions may lead to eruptions not only when they find their way to the surface, but also when they hit magma residing in the roots of volcanoes. At Eyjafjallajökull our observations show how primitive melts in an intrusive complex active since 1992 triggered an explosive eruption of trachyandesite.





**Figure 3 | Inferred sources of deformation from 'hydrostatic crack' models with variable opening, shown in map view, together with a cross-section of the summit area.** Opening values represent the mean opening of each patch from the posterior probability distribution, only plotted when this value is greater than the standard deviation of the distribution. Background maps give shaded topography, with ice caps in white. Black dots are earthquake epicentres in respective time periods. **a**, Initial sill opening until 28 February 2010. **b**, Continued sill opening 1–20 March. **c**, Inferred tilted dyke opening, also in

1–20 March period. **d**, Deflation source, modelled as a contracting sill, 14–22 April. **e**, Outline of all sources shown in **a–d**. Southern solid outline, 'sill 1' (see **f**) until end of February; northern solid outline and bar, continuing sill evolution (modelled as 'sill 2', see **f**) and dyke 1–20 March; dashed line, contracting sill 14–22 April. **f**, Schematic east–west cross-section across the summit area, with sources plotted at their best-fit depth (vertical exaggeration by a factor of 2). Grey shaded background indicates source depth uncertainties (95% confidence), which overlap.

## METHODS SUMMARY

GPS data were analysed using GAMIT/GLOBK<sup>24</sup>. InSAR images were formed with Doris software<sup>25</sup> using scripts from StaMPS<sup>17</sup>. Earthquake hypocentres were located using data acquired by the SIL system<sup>26</sup>. The chemical composition of eruptive products was determined using inductively coupled plasma optical emission spectrometry. In the joint inversion of the GPS and InSAR data sets, we solved for the opening of each element for a given hydrostatic overpressure, using a boundary

element model, assuming a density difference between the magma and surrounding rock of  $250 \text{ kg m}^{-3}$  and a traction-free interface<sup>27</sup>. Our model extends that of ref. 28 but our imposition of hydrostatic rather than uniform overpressure means that the overpressure decreases with depth. An exponential covariance function with variance of  $1 \text{ cm}^2$  for each interferometric pair was assumed. Our estimate of the model parameters accounts for 87% of the variance in the data and the results are supported by supplementary modelling of InSAR<sup>29</sup> and GPS data alone<sup>30</sup> that we carried out.



**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 14 May; accepted 5 October 2010.**

- Dzurisin, D. *Volcano Deformation* (Springer, 2007).
- Amelung, F., Jónsson, S., Zebker, H. & Segall, P. Widespread uplift and trapdoor faulting on Galápagos volcanoes observed with radar interferometry. *Nature* **407**, 993–996 (2000).
- Sigmundsson, F. *Iceland Geodynamics: Crustal Deformation and Divergent Plate Tectonics* (Springer, 2006).
- Pedersen, R. & Sigmundsson, F. InSAR based sill model links spatially offset areas of deformation and seismicity for the 1994 unrest episode at Eyjafjallajökull volcano, Iceland. *Geophys. Res. Lett.* **31**, L14610, doi:10.1029/2004GL020368 (2004).
- Pedersen, R. & Sigmundsson, F. Temporal development of the 1999 intrusive episode in the Eyjafjallajökull volcano, Iceland, derived from InSAR images. *Bull. Volcanol.* **68**, 377–393 (2006).
- Hooper, A., Pedersen, R. & Sigmundsson, F. in *The VOLUME Project — Volcanoes: Understanding Subsurface Mass Movement* (eds Bean, C. J. et al.) 13–24 (University College Dublin, 2009).
- Sturkell, E., Sigmundsson, F. & Einarsson, P. Recent unrest and magma movements at Eyjafjallajökull and Katla Volcanoes, Iceland. *J. Geophys. Res.* **108**, 2369, doi:10.1029/2001JB000917 (2003).
- Sturkell, E. et al. Katla and Eyjafjallajökull Volcanoes. *Dev. Quat. Sci.* **13**, 5–21 (2010).
- Hooper, A. A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches. *Geophys. Res. Lett.* **35**, L16302, doi:10.1029/2008GL034654 (2008).
- Oskarsson, N., Sigvaldason, G. E. & Steinthórsson, S. A dynamic model of rift zone petrogenesis and the regional petrology of Iceland. *J. Petrol.* **23**, 28–74 (1985).
- Larsen, G., Dugmore, A. J. & Newton, A. J. Geochemistry of historical-age silicic tephras in Iceland. *Holocene* **9**, 463–471 (1999).
- Hjaltadóttir, S. & Vogfjörð, K. Seismic evidence of magma transport in Eyjafjallajökull during 2009–2010. *Eos* **91**(Fall meet. suppl.) (abstract V21F-02).
- Gudmundsson, M. T. et al. The Eyjafjallajökull eruption in April–May 2010; course of events, ash generation and ash dispersal. *Eos* **91**(Fall meet. suppl.) (abstract V53F-01).
- Eineder, M. & Fritz, T. *TerraSAR-X Ground Segment Basic Product Specification Document TX-GS-DD-3302* (DLR, 2009).
- Okada, Y. Internal deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1992).
- Fialko, Y., Khazan, Y. & Simons, M. Deformation due to a pressurized horizontal circular crack in an elastic half-space, with applications to volcano geodesy. *Geophys. J. Int.* **146**, 181–190 (2001).
- Hooper, A., Segall, P. & Zebker, H. Persistent scatterer interferometric synthetic aperture radar for crustal deformation analysis, with application to Volcan Alcedo, Galapagos. *J. Geophys. Res.* **112**, B07407, doi:10.1029/2006JB004763 (2007).
- Hjaltadóttir, S., Vogfjörð, K. S. & Slunga R. *Seismic Signs of Magma Pathways Through the Crust in the Eyjafjallajökull Volcano, South Iceland*. (Icelandic Meteorological Office report VI 2009-13, 2009).
- Vogfjörð, K. S., Hjaltadóttir, S., Kjartansson, E. & Roberts, M. J. in *The VOLUME Project — Volcanoes: Understanding Subsurface Mass Movement* (eds Bean, C. J. et al.) 237–249 (University College Dublin, 2009).
- Taisne, B. & Jaupart, C. Dike propagation through layered rocks. *J. Geophys. Res.* **114**, B09203, doi:10.1029/2008JB006228 (2009).
- Sigmarrson O., Vlastelic, I. & Devidal, J. Trace-element variations reveal dynamic magma mixing during the 2010 eruption of Eyjafjallajökull, Iceland. *Eos* **91**(Fall meet. suppl.) (abstract V21F-04).
- Eichelberger, J. C., Izbekov, P. E. & Browne, B. L. Bulk chemical trends at arc volcanoes are not liquid lines of descent. *Lithos* **87**, 135–154 (2006).
- Marsh, B. A magmatic mush column Rosetta Stone: The McMurdo Dry Valleys of Antarctica. *Eos* **47**, 497–508 (2004).
- Hreinsdóttir, S. & Bennett, R. A. Active aseismic creep on the Alto Tiberina low-angle normal fault, Italy. *Geology* **37**, 683–686 (2009).
- Kampes, B., Hanssen, R. F. & Zbigniew, P. Radar interferometry with public domain tools. In *Proceedings 'Fringe 2003 Workshop'* (ESA SP-550, European Space Agency, 2004).
- Jakobsdóttir, S. S. Seismicity in Iceland: 1994–2007. *Jökull* **58**, 75–100 (2008).
- Hooper, A. et al. Is magma generated due to retreating ice caps, likely to erupt? Constraints on lower-crustal stress in Iceland from InSAR. In *Proceedings 'Fringe 2009 Workshop'* (ESA SP-677, European Space Agency, 2010).
- Yun, S., Segall, P. & Zebker, H. Constraints on magma chamber geometry at Sierra Negra volcano, Galapagos Islands, based on InSAR observations. *J. Volcanol. Geotherm. Res.* **150**, 232–243 (2006).
- Feigl, K. L. & Thurber, C. H. A method for modelling radar interferograms without phase unwrapping: application to the M 5 Fawnskin, California earthquake of 1992 December 4. *Geophys. J. Int.* **176**, 491–504 (2009).
- Árnadóttir, T., Jiang, W., Feigl, K. L., Geirsson, H. & Sturkell, E. Kinematic models of plate boundary deformation in southwest Iceland derived from GPS observations. *J. Geophys. Res.* **111**, B07402, doi:10.1029/2005JB003907 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge the efforts of Th. Jónsson, J. Hólmjárn, S. Steinthórsson, H. Ólafsson, Th. Ingvarsson and B. Brandsdóttir in making field operations run smoothly, as well as consultation and discussions with the staff of our institutes. We thank P. Segall and C. Wicks for comments on the manuscript, and NASA/GSFC for use of the MODIS image. Financial support for this work was received from the Icelandic Research Fund, the research fund at the University of Iceland, and the US National Science Foundation (grant EAR 1042103). Funds were received for hazard mitigation from the Icelandic government. We thank the Geodesy Laboratory at the University of Arizona for computing facilities and UNAVCO for technical support. GMT public domain software was used for some figures. TerraSAR-X data were provided by the German Space Agency under project number GEO0609.

**Author Contributions** F.S. coordinated the writing of the paper and the research it is based on; S.H. and F.S. supervised the installation of new semi-continuous GPS stations; A.H. formed the interferograms and modelled the joint geodetic dataset; S.H. analysed the GPS data and produced the GPS time series; T.Á. modelled the GPS displacements; A.H., F.S., K.L.F. and R.P. planned the InSAR data acquisitions; S.H., M.J.R., A.A., H.G., M.H., B.G.Ó., H.S., E.S., P.E. and F.S. did the GPS measurements; N.Ó. conducted the geochemical analysis; P.E. has been involved in monitoring the volcano since 1971; M.J.R., S.H., R.P., P.E., N.Ó. and F.S. were involved in daily monitoring of the eruptive activity; J.D. unwrapped the interferograms; K.L.F. formed additional interferograms and modelled them; A.H., R.P., S.H. and J.D. produced the figures; F.S., A.H., S.H., T.Á., R.P., N.Ó. P.E. and K.L.F. led the writing of the paper, with all authors commenting on and discussing its results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to F.S. (fs@hi.is).

## METHODS

**Seismology.** Earthquakes were located using the now national SIL (South Iceland Lowland) seismic network in Iceland operated by the Icelandic Meteorological Office (IMO). The network comprises 56 three-component digital seismic stations located throughout Iceland<sup>26,31,32</sup>. Eight SIL stations, sampling continuously at 100 Hz, are located within 50 km of the summit of Eyjafjallajökull. Over 2,800 earthquakes were used for analysis from the IMO catalogue, with each event interactively picked using the SIL velocity model<sup>18</sup>. Events were selected according to the following constraints: (1) waveform detections at three or more SIL stations; (2) sourced from within an area encompassing 63.2–63.8° N and 19.45–19.8° W; (3) occurring between 1 September 2009 and 30 April 2010; and (4) local magnitudes ( $M_L$ )  $\geq 0.8$ , denoting the minimum magnitude of completeness for the region. Mean, standard deviation, and maximum earthquake sizes were  $M_L$  1.2, 0.3, and 3.6, respectively. The absolute locational accuracy of the selected earthquakes was horizontally  $\pm 0.6$  km with a standard deviation of 0.01 km, and vertically  $\pm 1.5$  km with a standard deviation of 1.2 km. Focal depths ranged from less than 1 km to 23 km, with a mean concentration at a depth of 7.5 km (and standard deviation of 2.7 km). Note that the catalogue of earthquakes is incomplete as several hundred earthquakes from the pre-eruption swarms await manual analysis. The SIL network is operated by the IMO; for further details, see: <http://en.vedur.is/>.

**Radar interferometry.** We used radar data acquired by the TerraSAR-X satellite operated by the German Space Agency DLR<sup>14</sup> from both ascending track 132 and descending track 125. For track 132 the ground heading is 347.4° and the mean angle of incidence is 30.5°. For track 125 the ground heading is 190.8° and the mean angle of incidence is 37.4°. Interferograms were formed with Doris software<sup>25</sup> using scripts from StaMPS<sup>17</sup>. Precise orbits from DLR and the 25 m digital elevation model from the National Land Survey of Iceland were used to correct for the geometric component of the interferometric phase. The resulting phase values were unwrapped using a statistical cost flow algorithm<sup>33</sup> and then resampled using a quad-tree approach<sup>34</sup>.

**GPS.** The GPS stations around Eyjafjallajökull volcano are equipped with Trimble dual frequency GPS receivers and Chokering and Zephyr geodetic antennas. We analysed data from these GPS stations using GAMIT software version 10.3 (refs 24 and 35). In addition to data from continuous GPS stations in Iceland, we analysed data from over 150 global reference stations of which 35 were used to determine the ITRF05-Eurasia fixed reference frame<sup>36,37</sup>. We applied the ocean-loading model FES2004 and used the IGS absolute antenna phase centre models for both satellite and ground-based antennas. We used GLOBK software<sup>38</sup> to estimate the daily positions for the GPS stations. We estimated linear, annual and semi annual correction terms using data from THEY spanning 2002.0 to 2009.0 (ref. 39) to detrend all time series. Annual and semi annual terms for THEY were used for SKOG and STE2.

**Modelling.** Our preferred models are found by joint inversion of GPS and InSAR data; their details are listed in Supplementary Table 1. Two approaches were applied, the first using only simple geometries, with penny-shaped cracks<sup>16</sup> representing sills and a rectangular dislocation representing a dyke<sup>15</sup>. In our second modelling approach, we divided potential sources into multiple rectangular elements and solved for the opening of each element for a given hydrostatic overpressure, using a boundary elements approach. We assumed a density difference between the magma and surrounding rock of 250 kg m<sup>-3</sup> and a traction-free interface<sup>27</sup>. Our model extends that of ref. 28, but our imposition of hydrostatic rather than uniform overpressure means that the overpressure decreases with depth. An elastic halfspace was assumed in both approaches, with Poisson's ratio of 0.27 and a shear modulus of 30 GPa. We applied Markov chain Monte Carlo sampling<sup>17</sup> to build the probability distribution of the model parameters, assuming a uniform prior probability. An exponential covariance function with variance of 1 cm<sup>2</sup> and unknown decay constant was assumed for each interferogram, with the decay constants estimated during the inversion. Observations, predicted deformation from the maximum likelihood model, and residuals between observations and predictions for the two modelling approaches used are shown in Supplementary Figs 4 and 5. The 'hydrostatic cracks' model accounts for 87% of the variance in the data.

Complementary initial modelling was carried out with GPS data separately. The GPS time series were divided into several time intervals (1 Jan – 1 March, 1–20 March, 20 March – 13 April, 13–16 April, and 16–24 April) and the displacements modelled for each, assuming planar constant opening mode dislocations embedded in elastic half-space<sup>15</sup>. We assumed a Poisson's ratio of 0.25 and a shear modulus of 30 GPa. Nonlinear optimization was used to minimize the weighted residual sum of squares (WRSS) and the model fit judged by calculating a normalized  $\chi^2_v$  ( $=\text{WRSS}/(N-m)$ ) (ref. 30). We estimate 95% uncertainties in the model values using a bootstrap approach. For the bootstrap calculation, we sample the original data at random to generate a new data vector where data from one station may appear more than once, and data from another station may not be included. The GPS displacements in January and February are well fitted by a sill located at ~6 km depth, with a volume increase of  $18 \times 10^6$  m<sup>3</sup> ( $\chi^2_v = 1.8$ ). Bootstrap calculations

give a depth range for the sill of 4–8 km and a volume change of  $(13-30) \times 10^6$  m<sup>3</sup>. These bounds agree with the preferred model, although they give a larger range; this is expected, since there are many more InSAR data than GPS vectors used in the joint inversion, and the methods for estimating the model bounds are different. The GPS displacements leading up to the flank eruption are more difficult to fit with simple planar dislocations. Continued expansion of the sill found for the first time interval (volume increase of  $12 \times 10^6$  m<sup>3</sup>), accompanied by a dyke extending to the surface, has the best fit of the models tested ( $\chi^2_v = 35$ ). The total volume increase of the sill and dyke sources from 1 January to March 20 is about  $50 \times 10^6$  m<sup>3</sup>. Little deformation is observed during 20 March to 13 April, and models for that time interval have not been extensively tested. Following the onset of the second eruption, the GPS data indicate a contracting source at the summit of Eyjafjallajökull, located at 3–4 km depth, with decreasing volume change ( $14.6 \times 10^6$  m<sup>3</sup> during 13–16 April, and  $8 \times 10^6$  m<sup>3</sup> for 16–24 April). GPS displacements during the time interval spanned by the InSAR data (11–22 April) can be fitted by a single contracting sill at ~3 km depth with a volume decrease of  $21 \times 10^6$  m<sup>3</sup> ( $\chi^2_v = 4$ ). Bootstrap calculations give a depth range of 1–5 km and a volume decrease of  $(5-44) \times 10^6$  m<sup>3</sup>. Again, the range of model parameters is much larger than estimated for the preferred model, as explained above. The InSAR data were also considered separately, with the general inversion of phase technique<sup>29</sup>. It works directly on the wrapped phase, as shown in Fig. 2 and Supplementary Fig. 2. This algorithm minimizes a cost function that quantifies the misfit between observed and modelled values in terms of wrapped phase. Applying the general inversion of phase technique to an interferogram spanning the interval from 10 July 2009 through to 3 May 2010, we find an acceptable fit using a single sill at 6 km depth that inflates by  $(30-40) \times 10^6$  m<sup>3</sup>. Although this model neglects the time dependence of the deformation field and likely source complications, the inferred net volume change in the volcano roots is comparable to those found using the other modelling approaches.

**Geochemistry and petrology.** The chemical composition of eruptive products from the two eruptions were analysed with a Spectro ICP-OES (inductively coupled plasma optical emission spectrometry) spectrograph at the University of Iceland. Results reported in Supplementary Table 2 are the average of duplicate analysis, on two representative samples from the 2010 flank and summit eruptions, respectively. Rock powder was fluxed with lithium metaborate, and then dissolved in a mixture of nitric, hydrochloric and oxalic acids. Reference samples USGS-BHVO and USGS QLO-1 were used. The samples from the flank eruption are olivine basalt, weakly alkalic with minor normative nepheline. They are similar to those that make up most of the Eyjafjöll mountain<sup>40,41</sup> (the lower part of the Eyjafjallajökull volcano). On the basis of microscope work, the erupted olivine basalt has a very low degree of crystallinity, less than 2 wt% of olivine, plagioclase and clinopyroxene. Unzoned microphenocrysts of these minerals are also found dispersed in the otherwise almost aphyric groundmass. Together, these features indicate short crustal residence time before eruption. Olivine and plagioclase occur in glomerophytic aggregates, indicating shallow pre-eruptive equilibration at pressures about or below 4 kbar (above 13 km depth) where olivine and plagioclase are the first minerals to form. At higher pressures, clinopyroxene would be the first phase to form, followed by olivine in the absence of plagioclase. The summit eruption produced trachyandesitic magma of mildly alkalic iron-rich composition (common in the summit area of the mountain<sup>40</sup>). This magma is also very aphyric, with less than 2 wt% crystallinity. The mineralogy reveals equilibrium with the observed euhedral feldspar, clinopyroxene, olivine and spinel microphenocrysts. The low crystallinity suggests that the pre-eruptive crustal residence time of the andesite-magma body was short. Large euhedral xenocrysts of olivine, plagioclase and to a lesser extent clinopyroxene of basaltic origin occur within the trachyandesite. These crystals may be remnants of minor recent intrusion of olivine basalt. Xenolithic fragments of diorite, the plutonic equivalent of the andesite, are found in the rock.

31. Bódvarsson, R., Rögnvaldsson, S., Th., Slunga, R. & Stefánsson, R. The SIL data acquisition and monitoring system. *Seismol. Res. Lett.* **67**, 35–46 (1996).
32. Jakobsdóttir, S. S., Guðmundsson, G. B. & Stefánsson, R. Seismicity in Iceland 1991–2000 monitored by the SIL seismic system. *Jökull* **51**, 87–94 (2002).
33. Hooper, A. A statistical-cost approach to unwrapping the phase of InSAR time series. In *Proceedings 'Fringe 2009 Workshop'* (ESAP-677, European Space Agency, 2010).
34. Jónsson, S., Zebker, H., Segall, P. & Amelung, F. Fault slip distribution of the Mw 7.2 Hector Mine earthquake estimated from satellite radar and GPS measurements. *Bull. Seismol. Soc. Am.* **92**, 1377–1389 (2002).
35. King, R. W. & Bock, Y. *GAMIT Reference Manual, GPS Analysis at MIT, Release 10.3* (Technical report, MIT, 2006).
36. Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B. & Boucher, C. ITRF2005: A new release of the International Terrestrial Reference Frame based on time series of station positions and earth orientation parameters. *J. Geophys. Res.* **112**, B09401, doi:10.1029/2007JB004949 (2007).
37. Hreinsdóttir, S. et al. A complex earthquake sequence captured by the continuous GPS network in SW Iceland. *Geophys. Res. Lett.* **36**, L12309, doi:10.1029/2009GL038391 (2009).

38. Herring, T. A., King, R. W. & McClusky, S. C. *GLOBK Reference Manual, Global Kalman Filter VLBI and GPS Analysis Program, Release 10.3* (Technical report, MIT, 2006).
39. Herring, T. MATLAB Tools for viewing GPS velocities and time series. *GPS Solut.* **7**, 194–199 (2003).
40. Loughlin, S. C. *The Evolution of the Eyjafjöll Volcanic System, Southern Iceland*. Ph.D. thesis, Univ. Durham (1996).
41. Jakobsson, S. P. Petrology of Recent basalts of the Eastern Volcanic Zone, Iceland. *Acta Naturalia Islandica* **26**, 1–103 (1979).



# The trophic fingerprint of marine fisheries

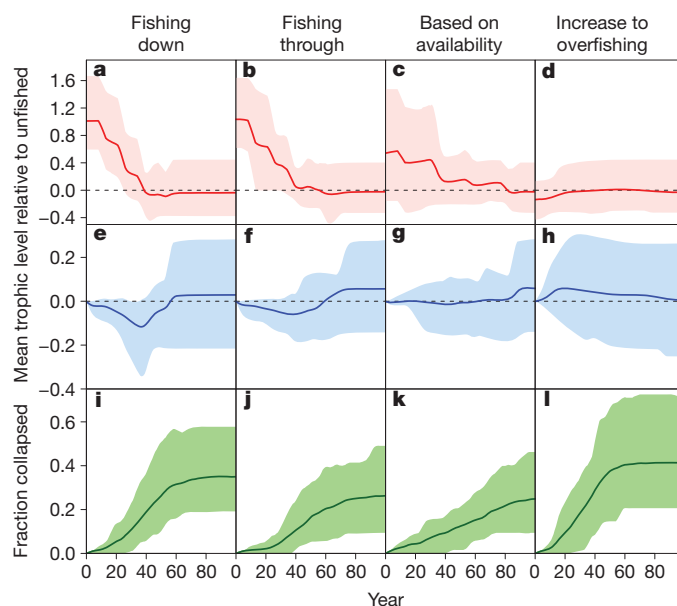
Trevor A. Branch<sup>1</sup>, Reg Watson<sup>2</sup>, Elizabeth A. Fulton<sup>3</sup>, Simon Jennings<sup>4,5</sup>, Carey R. McGilliard<sup>1</sup>, Grace T. Pablico<sup>2</sup>, Daniel Ricard<sup>6</sup> & Sean R. Tracey<sup>7</sup>

Biodiversity indicators provide a vital window on the state of the planet, guiding policy development and management<sup>1,2</sup>. The most widely adopted marine indicator is mean trophic level (MTL) from catches, intended to detect shifts from high-trophic-level predators to low-trophic-level invertebrates and plankton-feeders<sup>3–5</sup>. This indicator underpins reported trends in human impacts, declining when predators collapse (“fishing down marine food webs”) and when low-trophic-level fisheries expand (“fishing through marine food webs”)<sup>6</sup>. The assumption is that catch MTL measures changes in ecosystem MTL and biodiversity<sup>2,5</sup>. Here we combine model predictions with global assessments of MTL from catches, trawl surveys and fisheries stock assessments<sup>7</sup> and find that catch MTL does not reliably predict changes in marine ecosystems. Instead, catch MTL trends often diverge from ecosystem MTL trends obtained from surveys and assessments. In contrast to previous findings of rapid declines in catch MTL<sup>3</sup>, we observe recent increases in catch, survey and assessment MTL. However, catches from most trophic levels are rising, which can intensify fishery collapses even when MTL trends are stable or increasing. To detect fishing impacts on marine biodiversity, we recommend greater efforts to measure true abundance trends for marine species, especially those most vulnerable to fishing.

Adoption of an ecosystem approach to fisheries requires managers to conserve marine biodiversity, not just focus on fished stocks<sup>8</sup>. Biodiversity indicators are used to assess the impacts of fishing and the effectiveness of management, and thus guide the development of future policies<sup>9–12</sup>. The most widely used indicator, catch MTL, measures shifts in reported catches from high-trophic-level predators such as cod to low-trophic-level species such as filter-feeding oysters and small herbivorous fish<sup>3,13</sup>. In 1998, catch MTL was reported to be declining at an alarming 0.1 units per decade (“fishing down marine food webs”), and was interpreted to result from broad reductions in top predator biomass<sup>3–5</sup>. Catch MTL was the primary marine index chosen by the Convention on Biological Diversity to measure global biodiversity, and has been applied widely to report on the state of the marine environment<sup>1–5,9–12</sup>.

Catch MTL is interpreted to track changes in the underlying ecosystem<sup>3–5,14</sup>, but its usefulness as an indicator has been questioned because catches are influenced by changes in economics, management, fishing technology and targeting patterns<sup>6,15–20</sup>. Here we conducted the first large-scale test of whether catch MTL is a good indicator of ecosystem MTL, marine biodiversity and ecosystem status. We identified four main patterns of fisheries development and modelled their influence on MTL, and then compared these theoretical predictions with estimates of MTL from global compilations of catches, long-term trawl surveys, and fisheries stock assessments<sup>7</sup>, addressing three key questions: (1) whether catch MTL is positively correlated with ecosystem MTL, (2) what is the global MTL trend based on data from different sources, and (3) whether trends in MTL are informative about trends in marine ecosystem status.

We compiled ecosystem models<sup>21</sup> from 25 different ecosystems around the world, and simulated four main scenarios to examine the theoretical relation between catch MTL and ecosystem MTL (Fig. 1). The four scenarios were ‘fishing down’<sup>3</sup>, as already outlined, ‘fishing through’<sup>6</sup>, in which sequential expansion of low-trophic-level fisheries rather than collapses of top predators drives MTL, ‘based on availability’<sup>19</sup>, in which easily accessible species with high biomass are targeted first before expanding to less-accessible stocks with lower yields, and ‘increase to overfishing’, in which all species are fished with growing intensity over time until depleted. The simulations show that ‘fishing down’ and ‘fishing through’ both produce declining trends in catch MTL, but that ‘fishing down’ results in greater initial declines in ecosystem MTL, and more collapsed species than does ‘fishing through’. These scenarios predict that, at the end of the



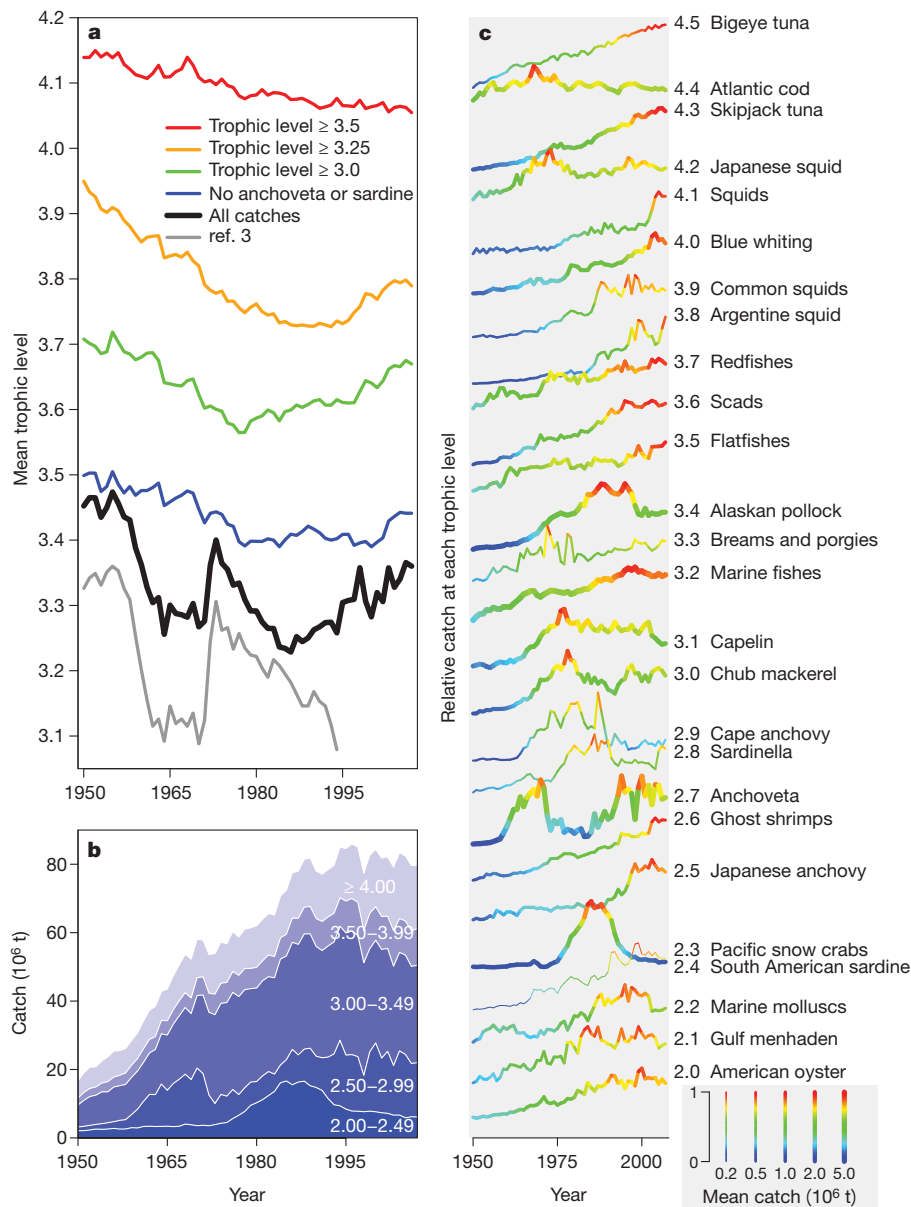
**Figure 1 | Changes in MTL relative to unfished ecosystem MTL.** Red, catches; blue, ecosystem biomass; green, the corresponding fraction of groups that are collapsed. Each panel shows the mean (solid line) and confidence intervals (10th and 90th, shading) of models from 25 ecosystems, for 100 years since the modelled start of fishery development. The scenarios are as follows. **a, e, i**, ‘Fishing down’: fishing top predators to depletion before sequentially switching to and depleting lower and lower trophic level groups. **b, f, j**, ‘Fishing through’: maintaining high catches of top predators while sequentially adding species at lower and lower trophic levels. **c, g, k**, ‘Based on availability’: targeting the most abundant and accessible taxa first before shifting to less-abundant and harder-to-access taxa. **d, h, l**, ‘Increase to overfishing’: expanding fishing mortality on all fished groups over time to twice the sustainable level for each group.

<sup>1</sup>School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, Washington 98195-5020, USA. <sup>2</sup>Sea Around Us Project, Fisheries Centre, University of British Columbia, Vancouver, V6T 1Z4, Canada. <sup>3</sup>CSIRO Wealth from Oceans, GPO Box 1538, Hobart, Tasmania 7001, Australia. <sup>4</sup>Centre for Environment, Fisheries and Aquaculture Science, Lowestoft NR33 0HT, UK. <sup>5</sup>School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. <sup>6</sup>Biology Department, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada. <sup>7</sup>Marine Research Laboratories, Tasmanian Aquaculture and Fisheries Institute, University of Tasmania, Private Bag 49, Hobart, Tasmania 7001, Australia.

simulations, most species are depleted (and many are collapsed to less than 10% of unexploited biomass), but MTL has returned to values observed in unexploited systems, because species across all trophic levels are equally depleted. More variability is observed in outcomes from the 'based on availability' scenario, which generally predicted declines in catch MTL, but less change in ecosystem MTL. Finally, the 'increase to overfishing' scenario hardly influenced catch and ecosystem MTL, but resulted in many collapsed species. These results (Fig. 1) are averaged over all models, and obscure substantial differences observed in particular models (Supplementary Figs 1–4). Overall, catch and ecosystem MTL were negatively correlated in many ecosystem models (35–38% of all models) in the 'fishing down', 'fishing through', and 'based on availability' scenarios, but usually positively correlated for the 'increase to overfishing' scenario and for

additional scenarios in which fishing was applied evenly across all species (Supplementary Figs 5–10). Importantly, this shows that when fishing disproportionately affects one part of the food web, the relation between catch MTL and ecosystem MTL often breaks down, but when fishing similarly affects all species, catches act as a representative sample of ecosystem changes.

We calculated catch MTL from global fishery landings, finding substantially different values and trends to those reported in ref. 3 (Fig. 2a). In particular, catch MTL has not declined steeply since the 1970s, but initially declined and then increased from the mid-1980s. Other recent publications reporting similar trends<sup>2,22,23</sup> have not explained why their results differ from those in ref. 3. We discovered that these differences arose from updates to the main source of trophic level estimates, FishBase<sup>24</sup>, and not from changes in relative catches



**Figure 2 | Trends in MTL from global marine catches.** **a**, Catch MTL in ref. 3. (grey), compared with catch MTL calculated from the most recent data (black), calculated after excluding anchoveta and South American sardine (blue), and calculated after excluding all species below trophic level 3.0 (green), 3.25 (orange), and 3.5 (red). **b**, Total catches divided into 0.5 trophic level bins. **c**, Relative catch trends divided into 0.1 trophic level bins, with the most

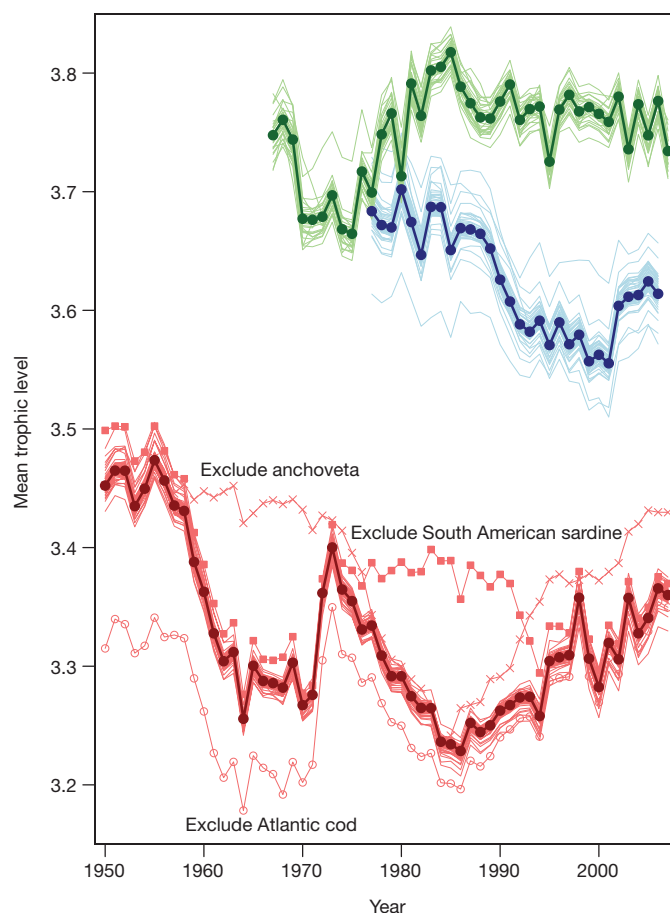
dominant taxon in each bin listed on the right (when summed, these taxa account for 50% of global catches). The legend for **c** at the bottom right explains that line colours are graded from zero (deep blue) to maximum relative catch (red) within each bin, while line width is proportional to average annual catch, metric tonnes.

among species. One key change was increasing the trophic level estimate of anchoveta from 2.2 to 2.7, which markedly altered the global catch MTL trend, and highlights the sensitivity of catch MTL trends to uncertainty in trophic level estimates (for more details see Supplementary Materials and Supplementary Figs 12–14).

In addition to anchoveta, global catch MTL trends are affected by other highly fluctuating stocks of small pelagic fishes. Dips and recoveries in catch MTL in the 1960s and 1980s were caused by the respective rapid development and collapse of anchoveta and sardine fisheries, which fluctuate in response to climate and fishing and are often out of phase with each other<sup>25</sup>. Catch MTL is much smoother over time when recalculated without these two species (Fig. 2a). Examining species grouped by 0.1 trophic level bins (Fig. 2c) reveals that catches of small pelagic species peaked at various times from the 1960s to the present<sup>25</sup>. Consequently, trends differ considerably when small pelagics are excluded by re-estimating catch MTL from groups with trophic levels above 3.0, 3.25 or 3.5 (ref. 5) (Fig. 2a). Declining trends in catch MTL within the remaining higher-trophic-level groups are driven by the collapse in Atlantic cod catches since the 1960s; removing Atlantic cod results in increasing catch MTL trends for groups above trophic levels 3.0, 3.25 and 3.5 (Supplementary Fig. 14). However, although Atlantic cod catches declined, catches of most other high-trophic-level predators expanded over time (Fig. 2c), while global catches increased until the mid-1980s and then levelled off<sup>23,26,27</sup> (Fig. 2b). Overall, fishing pressure has expanded at all levels of marine food webs, similar to our model scenario “increase to overfishing”.

Ecosystem MTL estimates were calculated in two ways: survey MTL from biomass estimates from 29 long-term trawl surveys, and assessment MTL from biomass estimates of 242 fisheries stock assessments. Trawl surveys offer consistent time series of ecosystem biomass, whereas assessments combine information from multiple sources to estimate biomass trends, focusing on important commercial stocks. Survey MTL is affected by catchability differences among species, and both survey MTL and assessment MTL are dependent on the selection of species that are surveyed or assessed, but both sources provide MTL estimates that can be used to measure ecosystem changes directly. We found that survey MTL and assessment MTL were higher than catch MTL (Fig. 3), reflecting the greater focus of surveys and stock assessments on bottom-dwelling high-trophic-level fish species that account for only a moderate proportion of total catch weight. Survey MTL initially declined, but is now higher than in the 1970s, whereas assessment MTL declined until the 1990s before recovering to within 0.05 units of the start value. Catch MTL was not positively correlated with ecosystem MTL. When all data are combined, catch MTL was negatively correlated with both survey MTL (Pearson correlation  $r = -0.55$ ) and assessment MTL ( $r = -0.31$ ) (Fig. 3); when restricted to a common set of stocks, catch MTL was also negatively correlated with assessment MTL ( $r = -0.41$ ) (Supplementary Fig. 18). These results indicate that catch MTL does not track changes in ecosystem MTL.

We also compared catch, survey and assessment MTL in individual ecosystems, finding that catch MTL is negatively correlated with survey MTL for 13 of 29 surveys, and negatively correlated with assessment MTL in 4 of 9 ecosystems. Three examples demonstrate these differences. In the Gulf of Alaska, catch and assessment MTL are dominated by Alaskan pollock and failed to capture the well-documented regime shift from low-trophic-level shrimp and crabs to high-trophic-level fish in the late 1970s<sup>28,29</sup>, but the Gulf of Alaska small-mesh shrimp survey did detect this shift, increasing 0.8 units (Fig. 4b, survey 3 in red). Conversely, the early-1980s collapse of cod and shift to invertebrates in eastern Canada (Fig. 4g, h) is captured by dramatic declines in catch MTL, but hardly visible in trawl surveys in the region, which lacked invertebrate data. Finally, in the Gulf of Thailand, where almost all fished species collapsed and survey MTL declined<sup>30</sup>, catch MTL increased continuously (Fig. 4m). The Gulf of Thailand pattern resulted

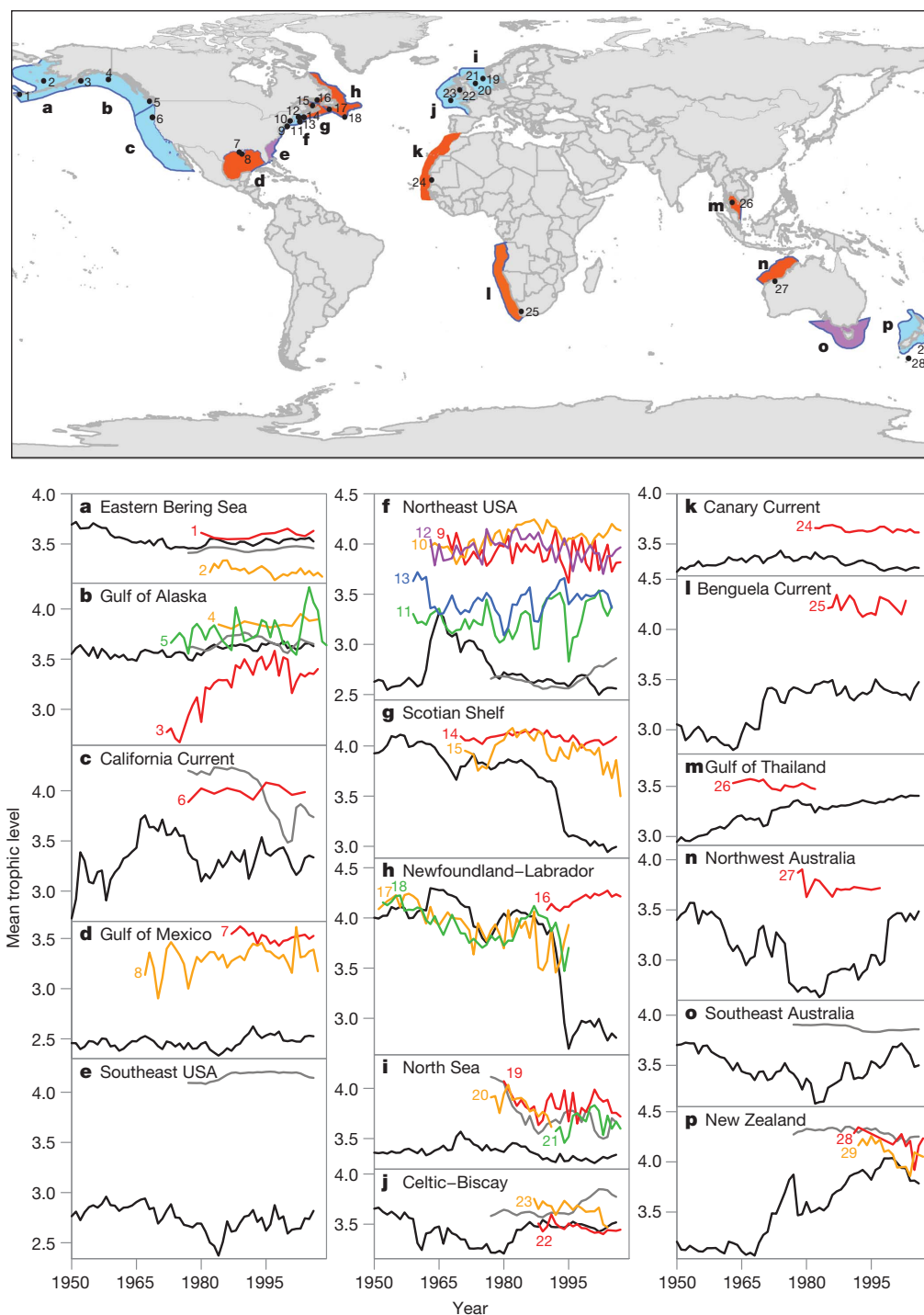


**Figure 3 | Measured MTL.** Thick lines show MTL from long-term trawl surveys (green), fisheries stock assessments (blue) and global catches (red). Faint lines show the effect of jack-knifing—excluding one unit at a time from the analysis and recalculating the respective trend. The exclusion of anchoveta (crosses), South American sardine (small squares), and Atlantic cod (open circles) substantially influenced the catch MTL time series.

from fishery development similar to the ‘based on availability’<sup>19</sup> scenario: fisheries first targeted the most accessible species yielding the highest revenue—mussels, shrimps and small fish—before expanding to high-trophic-level fish.

Global fisheries are at a crucial turning point, with high fishing pressure throughout marine food webs being offset in some regions by rebuilding efforts<sup>7</sup>. To measure the successes and failures of management, it is important for biodiversity indicators to track fishing impacts. Indicators such as catch MTL use readily available data and are quick and easy to calculate, but without improvement are ineffective measures of trends in biodiversity. Our theoretical models and empirical comparisons of catch MTL with ecosystem MTL suggest that catch MTL does not reliably measure the magnitude of fishing impacts or the rate at which marine ecosystems are being altered by fishing. Instead, we recommend a greater emphasis on measuring and reporting changes in marine biodiversity by tracking trends in abundance relative to reference points for conservation and sustainable use. To target limited resources in the best way, we should focus on assessing species vulnerable to fishing that are not currently assessed, and on developing and expanding trend-detection methods that can be applied more widely, particularly to countries with few resources for science and assessment. Through such efforts we can better detect and convey the true impact of fisheries on marine biodiversity.





**Figure 4 | MTL for each Large Marine Ecosystem.** The MTL is shown for each Large Marine Ecosystem from catches (black lines), assessments (grey lines) and surveys (colours). The map shows the location of each Large Marine

Ecosystem, highlighting those with data from all three sources (blue), from catches and surveys (red), and from catches and assessments (purple). Numbers on the map reflect the approximate centre of each survey.

## METHODS SUMMARY

Each taxon in the analysis was assigned a diet-based fractional trophic level, mostly from the online database FishBase<sup>24</sup>. Primary producers are trophic level one by definition, and were not included in our analyses; herbivores and filter feeders are trophic level two; and omnivores and carnivores are at higher trophic levels. MTL is the catch- or biomass-weighted average of trophic levels of taxa recorded in a particular year. Ecopath with Ecosim models<sup>21</sup> were compiled from well-documented sources and run for 100 years with zero catch to reach unfished states, and then four main scenarios of fishery development (fishing down<sup>3</sup>, fishing through<sup>6</sup>, based on availability<sup>19</sup>, and increase to overfishing) were applied during years 101 to 200. Global catch data were obtained from the United Nations Food and Agriculture Organization (FAO), while catch data for individual Large Marine

Ecosystems came from the Sea Around Us Project of the University of British Columbia; trends in catch MTL from these two sources are nearly identical. Long-term scientific trawl surveys from 15 Large Marine Ecosystems provide biomass estimates for regularly recorded taxa, and were obtained from a variety of sources. Biomass estimates for individual taxa were typically not corrected for differential catchability among taxa; furthermore, invertebrate biomass estimates were seldom included in the provided data. MTL time series from individual surveys were combined into a single global time series using a linear mixed effects model with 'Large Marine Ecosystem' modelled as a random effect. Stock assessment biomass values were obtained from the RAM Legacy database; total biomass was preferentially used in the analysis unless spawning biomass was the only time series available. Pearson correlations ( $r$ ) were used to assess whether MTL followed

the same trends in catches, surveys, and assessments, with statistical significance assessed after accounting for autocorrelation within time series.

Received 30 July; accepted 21 September 2010.

- Secretariat of the Convention on Biological Diversity. *Global Biodiversity Outlook 2* (Secretariat of the Convention on Biological Diversity, 2006).
- Butchart, S. H. M. *et al.* Global biodiversity: indicators of recent declines. *Science* **328**, 1164–1168 (2010).
- Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. J. Fishing down marine food webs. *Science* **279**, 860–863 (1998).
- Pauly, D. & Palomares, M.-L. Fishing down marine food web: it is far more pervasive than we thought. *Bull. Mar. Sci.* **76**, 197–211 (2005).
- Pauly, D. & Watson, R. Background and interpretation of the 'Marine Trophic Index' as a measure of biodiversity. *Phil. Trans. R. Soc. B* **360**, 415–423 (2005).
- Essington, T. E., Beaudreau, A. H. & Wiedenmann, J. Fishing through marine food webs. *Proc. Natl Acad. Sci. USA* **103**, 3171–3175 (2006).
- Worm, B. *et al.* Rebuilding global fisheries. *Science* **325**, 578–585 (2009).
- Pikitch, E. K. *et al.* Ecosystem-based fishery management. *Science* **305**, 346–347 (2004).
- Rochet, M.-J. & Trenkel, V. M. Which community indicators can measure the impact of fishing? A review and proposals. *Can. J. Fish. Aquat. Sci.* **60**, 86–99 (2003).
- Fulton, E. A., Smith, A. D. M. & Punt, A. E. Which ecological indicators can robustly detect effects of fishing? *ICES J. Mar. Sci.* **62**, 540–551 (2005).
- Shin, Y.-J. *et al.* Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 2. Setting the scene. *ICES J. Mar. Sci.* **67**, 692–716 (2010).
- Shannon, L. J., Coll, M. & Neira, S. Exploring the dynamics of ecological indicators using food web models fitted to time series of abundance and catch data. *Ecol. Indic.* **9**, 1078–1095 (2009).
- Yang, J. A tentative analysis of the trophic levels of North Sea fish. *Mar. Ecol. Prog. Ser.* **7**, 247–252 (1982).
- Pauly, D., Watson, R. & Alder, J. Global trends in world fisheries: impacts on marine ecosystems and food security. *Phil. Trans. R. Soc. B* **360**, 5–12 (2005).
- de Mutsert, K., Cowan, J. H. Jr, Essington, T. E. & Hilborn, R. Reanalyses of Gulf of Mexico fisheries data: landings can be misleading in assessments of fisheries and fisheries ecosystems. *Proc. Natl Acad. Sci. USA* **105**, 2740–2744 (2008).
- Caddy, J. F. & Garibaldi, L. Apparent changes in the trophic composition of world marine harvests: the perspective from the FAO capture database. *Ocean Coast. Manag.* **43**, 615–655 (2000).
- Caddy, J. F., Csirke, J., Garcia, S. M. & Grainger, R. J. R. How pervasive is "Fishing down marine food webs"? *Science* **282**, 1383a (1998).
- Powers, J. E. & Monk, M. H. Current and future use of indicators for ecosystem based fisheries management. *Mar. Policy* **34**, 723–727 (2010).
- Sethi, S. A., Branch, T. A. & Watson, R. Fishery development patterns are driven by profit but not trophic level. *Proc. Natl Acad. Sci. USA* **107**, 12163–12167 (2010).
- Litzow, M. A. & Urban, D. Fishing through (and up) Alaskan food webs. *Can. J. Fish. Aquat. Sci.* **66**, 201–211 (2009).
- Christensen, V. & Walters, C. J. Ecopath with Ecosim: methods, capabilities and limitations. *Ecol. Modell.* **172**, 109–139 (2004).
- Tacon, A. G. J., Metian, M., Turchini, G. M. & de Silva, S. S. Responsible aquaculture and trophic level implications to global fish supply. *Rev. Fish. Sci.* **18**, 94–105 (2010).
- Tacon, A. G. J. & Metian, M. Fishing for aquaculture: non-food use of small pelagic forage fish—a global perspective. *Rev. Fish. Sci.* **17**, 305–317 (2009).
- Froese, R. & Pauly, D. (eds) *FishBase* <www.fishbase.org> Version 03/2010 (2010).
- Chavez, F. P., Ryan, J., Lluch-Cota, S. E. & Niquen, C. M. From anchovies to sardines and back: multidecadal change in the Pacific Ocean. *Science* **299**, 217–221 (2003).
- Zeller, D. & Pauly, D. Good news, bad news: global fisheries discards are declining, but so are total catches. *Fish. Fish.* **6**, 156–159 (2005).
- Watson, R. & Pauly, D. Systematic distortions in world fisheries catch trends. *Nature* **414**, 534–536 (2001).
- Anderson, P. J. & Piatt, J. F. Community reorganization in the Gulf of Alaska following ocean climate regime shift. *Mar. Ecol. Prog. Ser.* **189**, 117–123 (1999).
- Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
- Christensen, V. Fishery-induced changes in a marine ecosystem: insight from models of the Gulf of Thailand. *J. Fish Biol.* **53** (Suppl. A), 128–142 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work arose out of a working group at the National Center for Ecological Analysis and Synthesis (NCEAS), funded by the University of California Santa Barbara, the US National Science Foundation (NSF), and the Moore Foundation. T.A.B. was additionally funded by the School of Aquatic and Fishery Sciences, University of Washington. The RAM Legacy stock assessment database was funded by the Canadian Natural Sciences and Engineering Research Council and the Canadian Foundation for Innovation; the Sea Around Us Project was funded by Pew Charitable Trusts. Additional collaborators and data contributors are acknowledged in the Supplementary Information.

**Author Contributions** T.A.B. designed the study, analysed the data, and wrote the paper; R.W. analysed the Sea Around Us Project catch data; E.A.F. designed and ran the ecosystem model analyses; S.J. analysed some trawl survey series; C.R.M. combined trawl survey data into a global time series; G.T.P. provided and calculated trophic level estimates; D.R. collated and analysed stock assessment data; and S.R.T. analysed the FAO and Sea Around Us Project catch data. All authors discussed the results and contributed to the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.A.B. ([tbranch@uw.edu](mailto:tbranch@uw.edu)).

# Climate-driven population divergence in sex-determining systems

Ido Pen<sup>1</sup>, Tobias Uller<sup>2</sup>, Barbara Feldmeyer<sup>1†</sup>, Anna Harts<sup>1</sup>, Geoffrey M. While<sup>3</sup> & Erik Wapstra<sup>3</sup>

Sex determination is a fundamental biological process, yet its mechanisms are remarkably diverse<sup>1,2</sup>. In vertebrates, sex can be determined by inherited genetic factors or by the temperature experienced during embryonic development<sup>2,3</sup>. However, the evolutionary causes of this diversity remain unknown. Here we show that live-bearing lizards at different climatic extremes of the species' distribution differ in their sex-determining mechanisms, with temperature-dependent sex determination in lowlands and genotypic sex determination in highlands. A theoretical model parameterized with field data accurately predicts this divergence in sex-determining systems and the consequence thereof for variation in cohort sex ratios among years. Furthermore, we show that divergent natural selection on sex determination across altitudes is caused by climatic effects on lizard life history and variation in the magnitude of between-year temperature fluctuations. Our results establish an adaptive explanation for intra-specific divergence in sex-determining systems driven by phenotypic plasticity and ecological selection, thereby providing a unifying framework for integrating the developmental, ecological and evolutionary basis for variation in vertebrate sex determination.

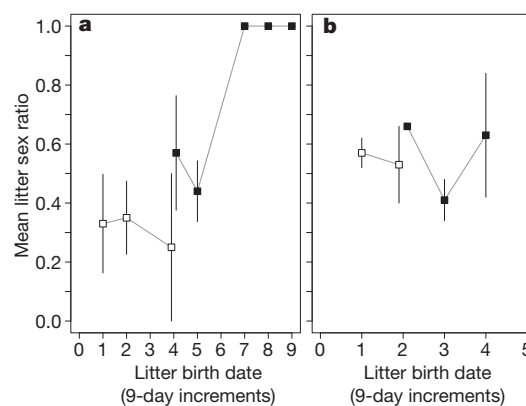
Vertebrates exhibit both genotypic (GSD) and temperature-dependent sex determination (TSD)<sup>1,2</sup>. The latter is particularly common in reptiles and both systems can co-occur within taxonomic families<sup>3</sup>. In addition, some species show elements of both genotypic and environmental sex determination within populations<sup>4,5</sup>. The causes of repeated evolutionary shifts between GSD and TSD and the origin and maintenance of mixed systems are two of the greatest unsolved problems in sex determination research<sup>1–8</sup>. The main reasons that diversity in reptilian sex determination has remained an enigma has been a failure empirically to link incubation temperature to ecological conditions promoting TSD and to establish theoretically that those conditions are sufficient to drive evolutionary shifts in sex-determining systems<sup>3,7</sup>. Here we provide both kinds of support using evolutionary models parameterized with field data to show how climatic effects on lizard life history generate evolutionary divergence in sex-determining systems via natural selection on sex ratios.

Environment-dependent sex determination can be favoured over genotypic sex determination when there are sex-specific fitness effects of environmental conditions experienced during or after the sex-determining period<sup>9</sup>. Temperature has a strong effect on the rate of embryonic development in ectotherm animals, with relatively cool conditions resulting in delayed birth or hatching. Sex differences in the fitness consequences of timing of birth could therefore favour integration of temperature-dependent developmental processes and gonad differentiation to ensure a match between offspring sex and birth date<sup>10,11</sup>. As a result, spatial or temporal variation in the strength of sex-specific selection on birth date, and therefore on TSD, may explain rapid evolutionary divergence in sex determination between populations or species<sup>10–12</sup>.

The snow skink, *Niveoscincus ocellatus*, is a small live-bearing lizard occurring along a 1,200-m altitudinal, and climatic, gradient from sea

level to highland regions throughout Tasmania<sup>13</sup>. Sex determination is affected by maternal basking opportunity in lowland skinks, analogous to temperature-dependent sex determination in egg-laying reptiles<sup>14</sup>. Thermal conditions representative of a cool year delays birth and result in an overproduction of male offspring whereas thermal conditions representative of warm years result in early birth and a small female bias (Fig. 1a). However, experimental manipulation of female thermal opportunity during gestation (a common garden experiment) reveals that sex determination in highland populations is not affected by temperature (Fig. 1b). This difference in sex-determining systems has consequences for sex ratios at the population level, with a negative correlation between the cohort sex ratio and annual temperature in lowland, but not highland, populations ( $r = -0.84$ ,  $P = 0.017$ ,  $N = 7$  and  $r = -0.20$ ,  $P = 0.65$ ,  $N = 7$ , respectively; slopes differ significantly between populations,  $F_{1,10} = 12.8$ ,  $P = 0.005$ ).

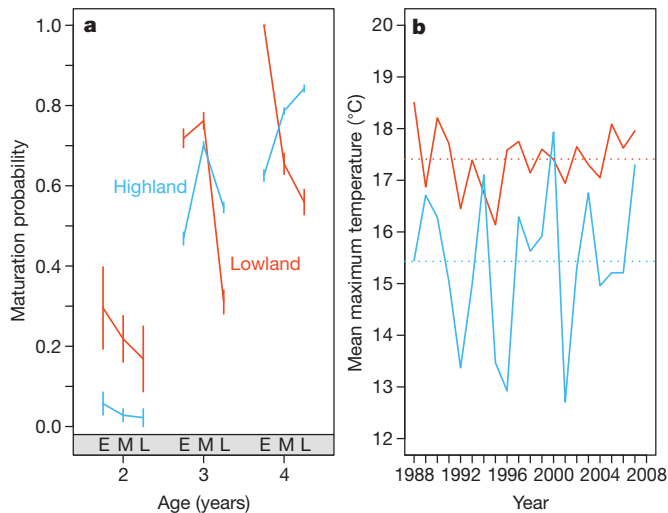
Earlier birth for females may be adaptive because birth date affects opportunity for growth until maturity, which is more important in female than in male snow skinks as a result of differences in selection on body size<sup>14–16</sup>. However, climatic conditions vary substantially across altitudes and the cooler conditions in highland regions induce several changes in lizard life history. High-altitude populations have a shorter activity season, more synchronized birth, slower growth and delayed age at maturity compared to lowland populations<sup>13,17</sup>. Birth date is therefore a relatively unimportant predictor of the onset of maturity and reproductive output at high altitudes (Fig. 2a). Specifically, at low altitudes early-born females have about 50% higher lifetime fitness than late-born females, whereas at high altitudes the effect of birth date on



**Figure 1 | Experimental effects of thermal conditions on sex ratio and birth date.** Sex ratio = male/(male + female). Poor thermal condition during gestation (filled squares) results in delayed birth compared to good thermal condition (open squares), with a corresponding significant effect on offspring sex in lowland (a) but not highland (b) females. Error bars are s.e.m. Logistic regression with the proportion of males as a dependent variable and treatment and birth date (measured in days from birth) as predictors: birth date for lowland population  $\chi^2 = 20.66$ ,  $P = 0.0001$ ,  $N_{\text{females}} = 13$ , 18 and for highland population,  $\chi^2 = 0.15$ ,  $P = 0.70$ ,  $N_{\text{females}} = 31$ , 24.

<sup>1</sup>Theoretical Biology Group, University of Groningen, PO Box 14, 9750 AA Haren, the Netherlands. <sup>2</sup>Edward Grey Institute, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. <sup>3</sup>School of Zoology, Private Bag 5, University of Tasmania, Hobart 7001, Tasmania, Australia. <sup>†</sup>Present address: Biodiversity and Climate Research Centre (BIK-F), Siesmayerstrasse 70A, D-60325 Frankfurt and Main, Germany.



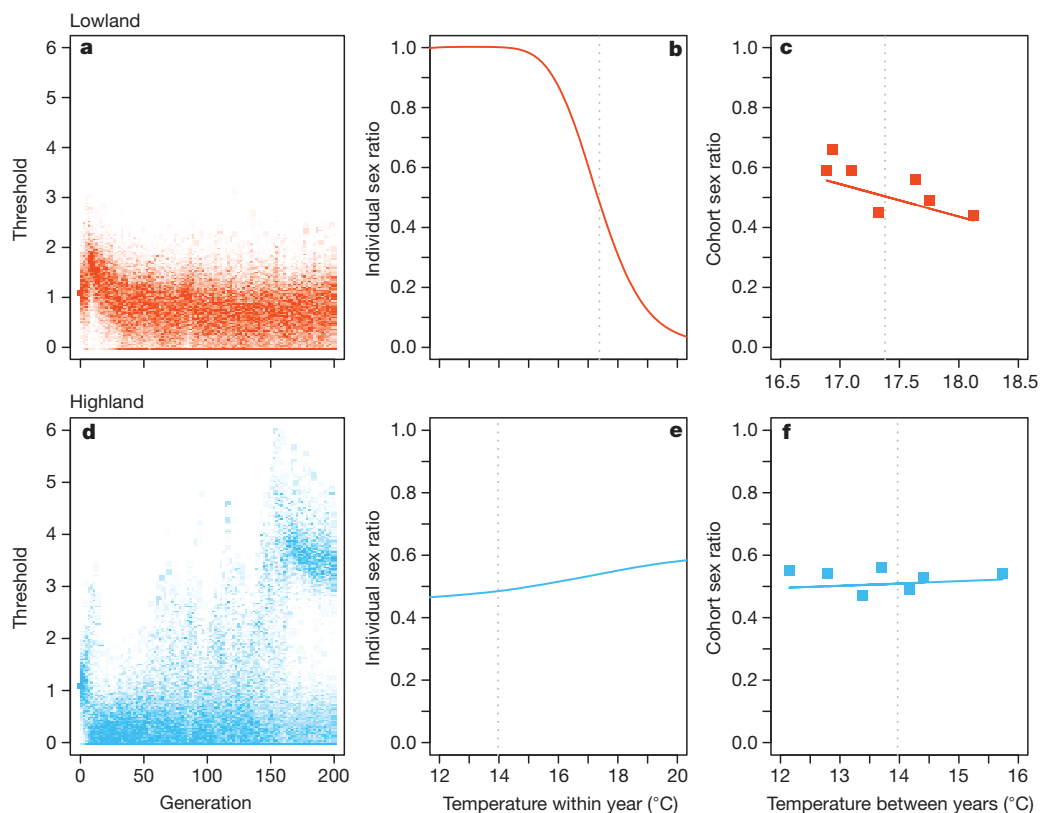


**Figure 2 | Life-history and temperature differences between lowland and highland populations of *N. ocellatus*.** **a**, Probability of maturation ( $\pm$ s.e.m.) at a given age for female offspring in relation to their timing of birth (E, early; M, intermediate; L, late) for lowland (red) and highland (blue) populations. Estimates based on field data from 2000–2007 (details provided in the Supplementary Information). **b**, Annual variation in mean maximum temperature experienced during the first half of gestation for lowland (red) and highland (blue) populations.

female fitness is greatly reduced (Fig. 2a; Supplementary Table 3). Furthermore, highland populations experience relatively high between-year variance in temperature (Fig. 2b), which could select for GSD because it prevents extreme sex ratios and therefore reduces variance in fitness across breeding attempts<sup>1,18–20</sup>.

To derive conditions under which the observed evolutionary divergence in sex determination in snow skinks could be favoured by natural selection, and to evaluate the relative importance of climate-induced changes in lizard life history and annual fluctuation in temperature, we constructed an individual-based simulation model based on a sex-determining mechanism recently proposed for lizards<sup>5</sup>. In this model, sex is determined by a threshold polymorphism involving four gene loci (see Supplementary Information for details). Each individual has a genetically determined temperature-dependent rate of regulatory gene expression, which needs to exceed a genetically determined threshold level to trigger male development (Supplementary Fig. 3). This allows evolutionary shifts in sex-determining systems via changes in the regulation of a developmental switch by genetic or environmental input. Both GSD and TSD can therefore be seen as emergent outcomes of selection for canalization of this switch, whereas ‘mixed systems’<sup>4,5</sup> occur when canalization is incomplete (Supplementary Information). We parameterized this model with empirical data from long-term studies of two populations at the climatic extremes of the species’ distribution and used sensitivity analyses to test whether climatic effects on life histories and the differences in the degree of between-year fluctuation in temperatures between altitudes were sufficient to explain the observed divergence in sex-determining systems. In addition, we calculated how well the temperatures experienced by individual females predicted their sex ratios to assess whether our model accurately captured the correlations observed in natural populations (see Methods and Supplementary Information for further details).

The model generated two primary results, both in close accordance with empirical data (Fig. 3). First, in simulations parameterized with data from the lowland population, sex determination evolved from pure GSD towards a system with a strong temperature effect (Fig. 3b). This



**Figure 3 | Evolutionary simulation results with genetic sex determination as ancestral state.** Upper panels, lowland parameter settings; lower panels, highland parameter settings. **a** and **d**, Population distributions of allelic values at threshold locus changing over time. We note branching in **d** for highland parameter settings, resulting in a novel sex-determining locus: males are ‘homozygous’ for alleles causing low thresholds and females ‘heterozygous’ for

low and high threshold alleles. **b** and **e**, Evolved average reaction norm for offspring sex ratio as a function of developmental temperature. The vertical dotted line is the average temperature experienced by natural populations. **c** and **f**, Predicted (from evolved reaction norm; line) and observed (natural populations; squares) cohort sex ratios for annual mean maximum temperature in the wild.

generated a significant negative correlation between the cohort sex ratio and average temperature during gestation that closely resembled data from our natural population (Fig. 3c). Second, in simulations parameterized with data from the highland population, sex chromosomes (W or Y) of the initial GSD system were either retained or, if lost, were replaced by a novel genetic element of major effect via disruptive selection on the threshold locus (Fig. 3d). Consequently, the model could generate evolutionary shifts from one sex chromosome system to another—including transitions between male and female heterogamety (Supplementary Information)—but it always produced a sex-determining system that generated average sex ratios that did not deviate substantially from equality, again in close accordance with our natural population (Fig. 3e, f). These results were robust with respect to starting settings, male versus female heterogamety, and linkage between genetic elements (Supplementary Information).

The population divergence in sex-determining systems could be explained by both the increased rate of female maturation with earlier birth date in lowland population and the higher magnitude of annual fluctuations in temperature in the highland population (Supplementary Fig. 4). Thus, a relatively long activity season favours an evolutionary shift from GSD to TSD in lowland populations, manifested in our model through the loss of genes of major effect and adaptive evolution of a sex ratio reaction norm and hence TSD. Conversely, a relatively cold and more variable climate reduces the activity season and delays maturity, which results in minor birth date effects on female age and size at maturity and causes disruptive selection on regulatory elements in sex-determining networks and the emergence of novel sex chromosomes. This model may also capture observed population or species divergence in sex-determining systems in fish<sup>10,12</sup> and thus may be generally applied to short-lived species.

Climate-driven population divergence in sex-determining systems emphasizes a creative role of phenotypic plasticity in evolution<sup>21</sup>. First, the effect of climate on lizard life history is largely a passive result of how thermal opportunity constrains activity patterns rather than an evolved adaptation<sup>22,23</sup>. However, such non-adaptive plasticity can apparently contribute to divergent selection on seasonal sex ratio adjustment and, hence, sex-determining mechanisms across species' distributions. Second, the observation that stressfully high or low temperatures have a causal effect on sex determination also in vertebrates with GSD<sup>5,24</sup> suggests that temperature-induced developmental plasticity can simultaneously expose variation in sex determination and cause novel selection on this variation, thereby greatly facilitating evolutionary divergence in sex-determining systems<sup>21,25</sup>. If so, transitions between sex-determining systems may only require minor secondary modifications in the regulation of gonad differentiation, suggesting substantial scope for interchangeability between genetic and environmental determinants of sex<sup>21</sup>.

## METHODS SUMMARY

All data are based on field studies of two intensively monitored populations at the climatic extremes of the species' distribution<sup>13,17,26</sup> and from the Bureau of Meteorology station situated close to our study sites. Females undergo gestation in the field and are brought into the laboratory just before birth to enable assessment of sex ratios and reproductive output<sup>26</sup>. The data were used to estimate survival, onset of maturity and reproductive output as a function of birth date to generate parameter estimates for the simulation model (see Supplementary Information). We used the mean daily maximum temperatures during the period of temperature-sensitivity of embryos as our index of thermal opportunity<sup>26,27</sup>.

To test directly the effect of thermal opportunity on sex determination we captured females early in gestation from areas adjacent to each of our main study sites and split them into two groups per population: extended basking conditions representative of warm years in lowland populations and limited basking conditions representative of cool years in highland populations (see ref. 14 for further detail).

Our simulation model is polygenic<sup>18</sup> and based on a dosage sex-determining mechanism recently proposed for lizards<sup>5</sup>. Sex is a threshold polymorphism determined by allelic values at four different loci (see Supplementary Information for details). We used daily temperatures from the past 20 years to calculate the long-term yearly mean ( $T_M$ ) and the annual variation ( $\sigma_B$ ) in temperature as well as the

within-year variation ( $\sigma_W$ ) in temperature. Each of 20 simulations started with 5,000 males and 5,000 females and the same values for reaction norm and threshold loci, with the age set to the minimum age at maturation. All results are from simulations run for 200,000 years.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 21 May; accepted 15 September 2010.**

**Published online 27 October 2010.**

1. Bull, J. J. *The Evolution of Sex Determining Systems* (Benjamin/Cummings Inc., 1983).
2. Valenzuela, N. & Lance, V. *Temperature-Dependent Sex Determination in Vertebrates* (Smithsonian Books, 2004).
3. Janzen, F. J. & Phillips, P. C. Exploring the evolution of environmental sex determination, especially in reptiles. *J. Evol. Biol.* **19**, 1775–1784 (2006).
4. Radder, R., Pike, D. A., Quinn, A. E. & Shine, R. Offspring sex in a lizard depends on egg size. *Curr. Biol.* **19**, 1–4 (2009).
5. Quinn, A. E. *et al.* Temperature sex reversal implies sex gene dosage in a reptile. *Science* **316**, 411–416 (2007).
6. Crews, D. & Bull, J. J. Mode and tempo in environmental sex determination in vertebrates. *Semin. Cell Dev. Biol.* **20**, 251–255 (2009).
7. Warner, D. A. & Shine, R. The adaptive significance of temperature-dependent sex determination in a reptile. *Nature* **451**, 566–569 (2008).
8. Uller, T., Pen, I., Wapstra, E., Beukeboom, L. W. & Komdeur, J. The evolution of sex ratios and sex-determining systems. *Trends Ecol. Evol.* **22**, 292–297 (2007).
9. Charnov, E. & Bull, J. J. When is sex environmentally determined? *Nature* **266**, 828–830 (1977).
10. Conover, D. O. Adaptive significance of temperature-dependent sex determination in a fish. *Am. Nat.* **123**, 297–313 (1984).
11. Warner, D. A., Uller, T. & Shine, R. Fitness effects of the timing of hatching may drive the evolution of temperature-dependent sex determination in short-lived lizards. *Evol. Ecol.* **23**, 281–294 (2009).
12. Conover, D. O. & Heins, S. W. Adaptive variation in environmental and genetic sex determination in a fish. *Nature* **326**, 496–498 (1987).
13. Wapstra, E. & Swain, R. Geographic and annual variation in life history traits in a temperate zone Australian skink. *J. Herpetol.* **35**, 194–203 (2001).
14. Wapstra, E. *et al.* Maternal basking behaviour determines offspring sex in a viviparous reptile. *Proc. R. Soc. Lond. B* **271**, S230–S232 (2004).
15. Olsson, M. *et al.* Sexual dimorphism in lizard body shape: the roles of sexual selection and fecundity selection. *Evolution* **56**, 1538–1542 (2002).
16. Atkins, N. *Parental Effects in Two Species of Viviparous Lizards: Niveoscincus microlepidotus and N. ocellatus*. PhD thesis, University of Tasmania (2007).
17. Wapstra, E., Swain, R. & O'Reilly, J. M. Geographic variation in age and size at maturity in a small Australian viviparous skink. *Copeia* **3**, 646–655 (2001).
18. Bulmer, M. G. & Bull, J. J. Models of polygenic sex determination and sex ratio control. *Evolution* **36**, 13–26 (1982).
19. van Dooren, T. J. M. & Leimar, O. The evolution of environmental and genetic sex determination in fluctuating environments. *Evolution* **57**, 2667–2677 (2003).
20. Leimar, O., Hammerstein, P. & van Dooren, T. J. M. A new perspective on developmental plasticity and the principles of adaptive morph determination. *Am. Nat.* **167**, 367–376 (2006).
21. West-Eberhard, M.-J. *Developmental Plasticity and Evolution* (Oxford University Press, 2003).
22. Adolph, S. C. & Porter, W. P. Growth, seasonality, and lizard life histories: age and size at maturity. *Oikos* **77**, 267–278 (1996).
23. Cadby, C. D. *et al.* Multi-scale approach to understanding climate effects on offspring size at birth and date of birth in a reptile. *Integr. Zool.* **5**, 164–175 (2010).
24. Ospina-Alvarez, N. & Pifferer, F. Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS ONE* **3**, E2837 (2008).
25. Uller, T. & Badyaev, A. V. Evolution of “determinants” in sex determination: a novel hypothesis for the origin of environmental contingencies in avian sex-bias. *Semin. Cell Dev. Biol.* **20**, 304–312 (2009).
26. Wapstra, E. *et al.* Climate effects on offspring sex ratio in a viviparous lizard. *J. Anim. Ecol.* **78**, 84–90 (2009).
27. Neaves, L. *et al.* Embryonic gonadal and sexual organ development in a small viviparous skink, *Niveoscincus ocellatus*. *J. Exp. Zool.* **305A**, 74–82 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Funding was provided by the Australian Research Council to E.W., T.U. and I.P. (DP0877948), by the Hermon Slade Foundation to E.W., T.U. and I.P., and by the Wenner-Gren Foundations to T.U.

**Author Contributions** T.U., I.P. and E.W. initiated, planned and coordinated the project; E.W. collected field and experimental data, assisted by G.M.W.; T.U., G.M.W. and I.P. analysed data and generated parameter estimates for the model; I.P., B.F., A.H. and T.U. constructed the model and analysed its outcome; T.U. and I.P. wrote the paper with input from all other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to I.P. (i.r.pen@rug.nl).

## METHODS

**Field procedures and data collection.** Between 2000/2001 and 2007/2008 approximately 90% of females from one lowland and one highland population of *N. ocellatus* were captured every year at the end of gestation, just before giving birth, resulting in a total of >1,500 females and >4,500 offspring. The taxonomic status of the populations as a single species and details on differences in life history traits have been described elsewhere<sup>13,17,28</sup>. Females were housed in cages until parturition, when all offspring were measured and sexed using hemipene eversion (repeatability >0.98 on the basis of animals followed to sexual maturity)<sup>26</sup>. Sex in this species is determined during the first half of gestation<sup>27</sup>. Offspring were released back into their population of origin randomly at 12 locations within each population. Paternity was assessed in a subset of litters using microsatellites<sup>16</sup>. The field data was used to estimate survival, onset of maturity, and reproductive output as a function of birth date, which were subsequently used as parameter estimates for the simulation model (see below; Supplementary Table 1).

**Common garden experiment.** Females captured early in gestation (before sex determination is completed<sup>27</sup>) from areas adjacent to each of our main study sites were split into two groups per population: extended basking conditions representative of warm years in lowland populations (10 h of basking per 24 h) and limited basking conditions representative of cool years in highland populations (4 h of basking per 24 h)<sup>13,14</sup>. At parturition, offspring were measured and sexed as for the natural populations. Sex-specific mortality can be ruled out because the number of offspring corresponded to the number of ovulated eggs assessed using palpation.

**Climate data.** Climatic data was obtained from Bureau of Meteorology stations situated close to our study sites. As a measure of the thermal conditions (basking opportunity) experienced by individual female skinks while gravid in the field we used the mean of daily maximum temperatures during gestation (first half of gestation, assigned as 1 October to 15 November in lowland and 15 October to 1 December in highland populations), which is an accurate determinant of the temperature experienced during sex determination<sup>26</sup>.

**Simulation model.** Our model is polygenic<sup>18</sup> and based on a dosage sex-determining system recently proposed for lizards<sup>5</sup>. Sex is a threshold polymorphism determined

by allelic values at four different loci (see Supplementary Information for details). On the basis of daily temperatures from the past 20 years (from each altitude) we calculated the long-term yearly mean ( $T_M$ ) and the annual variation ( $\sigma_B$ ) in temperature as well as the within-year variation ( $\sigma_W$ ) in temperature. In the model the yearly temperature ( $T_Y$ ) is calculated at each time step by drawing a value from a normal distribution with mean  $T_M$  and standard deviation  $\sigma_B$ .  $T_Y$  is further used to calculate female-specific thermal conditions ( $T_F$ ) by drawing a value from a normal distribution with mean  $T_Y$  and standard deviation  $\sigma_W$ . To facilitate model building, we divided each reproductive season into three categories: early, intermediate and late breeding (see Supplementary Information for further detail).

Data from our long-term study of two focal populations were used to estimate the minimum age at maturation, number of offspring, offspring and adult survival, and the probability of breeding at age  $t$  (Supplementary Information). Because age and body size do not influence male reproductive success in snow skinks<sup>15,16</sup>, we set the effect of birth date on male reproductive fitness to be zero. Each of 20 simulations started with 5,000 males and 5,000 females and the same values for reaction norm and threshold loci, and with the age set to the minimum age at maturation. The life history follows a simple structure (Supplementary Fig. 1). In brief, females mate with a randomly drawn male and produce a number of offspring according to her age drawn from a distribution of clutch sizes. The sex of the offspring is determined by the number of Z (or X) chromosomes, the reaction norm and threshold loci, and  $T_F$  (Supplementary Fig. 3). Offspring have a fixed probability of survival to the next year (survival is independent of birth date; Supplementary Information). Offspring that have reached the minimum age at maturation have a fixed age-specific probability of reproducing that depends on their timing of birth. At the end of each time step all individuals in the population age by one year and the cycle is restarted. All results are from simulations run for 200,000 years.

28. Melville, J. & Swain, R. Evolutionary relationships between morphology, performance and habitat openness in the lizard genus *Niveoscincus* (Scincidae: Lyosomaniidae). *Biol. J. Linn. Soc.* **70**, 667–680 (2000).



# A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria

Stephanie K. Aoki<sup>1</sup>, Elie J. Diner<sup>2</sup>, Claire t'Kint de Roodenbeke<sup>1</sup>, Brandt R. Burgess<sup>1</sup>, Stephen J. Poole<sup>1</sup>, Bruce A. Braaten<sup>1</sup>, Allison M. Jones<sup>1</sup>, Julia S. Webb<sup>1</sup>, Christopher S. Hayes<sup>1,2</sup>, Peggy A. Cotter<sup>1,2,\*†</sup> & David A. Low<sup>1,2\*</sup>

Bacteria have developed mechanisms to communicate and compete with one another in diverse environments<sup>1</sup>. A new form of intercellular communication, contact-dependent growth inhibition (CDI), was discovered recently in *Escherichia coli*<sup>2</sup>. CDI is mediated by the CdiB/CdiA two-partner secretion (TPS) system. CdiB facilitates secretion of the CdiA 'exoprotein' onto the cell surface. An additional small immunity protein (CdiI) protects CDI<sup>+</sup> cells from autoinhibition<sup>2,3</sup>. The mechanisms by which CDI blocks cell growth and by which CdiI counteracts this growth arrest are unknown. Moreover, the existence of CDI activity in other bacteria has not been explored. Here we show that the CDI growth inhibitory activity resides within the carboxy-terminal region of CdiA (CdiA-CT), and that CdiI binds and inactivates cognate CdiA-CT, but not heterologous CdiA-CT. Bioinformatic and experimental analyses show that multiple bacterial species encode functional CDI systems with high sequence variability in the CdiA-CT and CdiI coding regions. CdiA-CT heterogeneity implies that a range of toxic activities are used during CDI. Indeed, CdiA-CTs from uropathogenic *E. coli* and the plant pathogen *Dickeya dadantii* have different nuclease activities, each providing a distinct mechanism of growth inhibition. Finally, we show that bacteria lacking the CdiA-CT and CdiI coding regions are unable to compete with isogenic wild-type CDI<sup>+</sup> cells both in laboratory media and on a eukaryotic host. Taken together, these results suggest that CDI systems constitute an intricate immunity network with an important function in bacterial competition.

CDI was discovered in *E. coli* strain EC93, which inhibits the growth of other *E. coli* strains on direct cell-to-cell contact<sup>2</sup>. Epitope insertion mutagenesis revealed the importance of the CdiA-CT in CDI<sup>2</sup>. Genetic and antibody blocking experiments identified BamA, an essential protein required for outer membrane biogenesis, as the CDI receptor on target cells<sup>3</sup>. The inner membrane multidrug transporter AcrB may also have a function, because *acrB* mutants, like *bamA* mutants, are resistant to CDI<sup>3</sup>. For EC93-mediated CDI, growth inhibition coincides with dissipation of the proton motive force across the cytoplasmic membrane, decreased aerobic respiration and decreased ATP levels in the target cells<sup>4</sup>. EC93 is protected from autoinhibition by CdiI, which is encoded immediately downstream of *cdiA* (ref. 2). These data suggest that CdiA binds BamA and delivers a signal, possibly a CdiA-derived toxin, which then inhibits target cell growth. CdiI could confer immunity on cells by binding to the CdiA peptide or otherwise neutralizing the growth inhibitory signal (Supplementary Fig. 1a).

Previous complementation analysis indicated the presence of functional *cdiB* and *cdiA* homologues in uropathogenic *E. coli* (UPEC), but no *cdiI* homologue was identified<sup>2,5</sup>. Inspection of the *cdi* locus from *E. coli* UPEC 536 revealed a small open reading frame in the same relative location as, but lacking significant sequence identity to, *cdiI*<sub>EC93</sub>. Expression of this open reading frame (*cdiI*<sub>536</sub>) protected

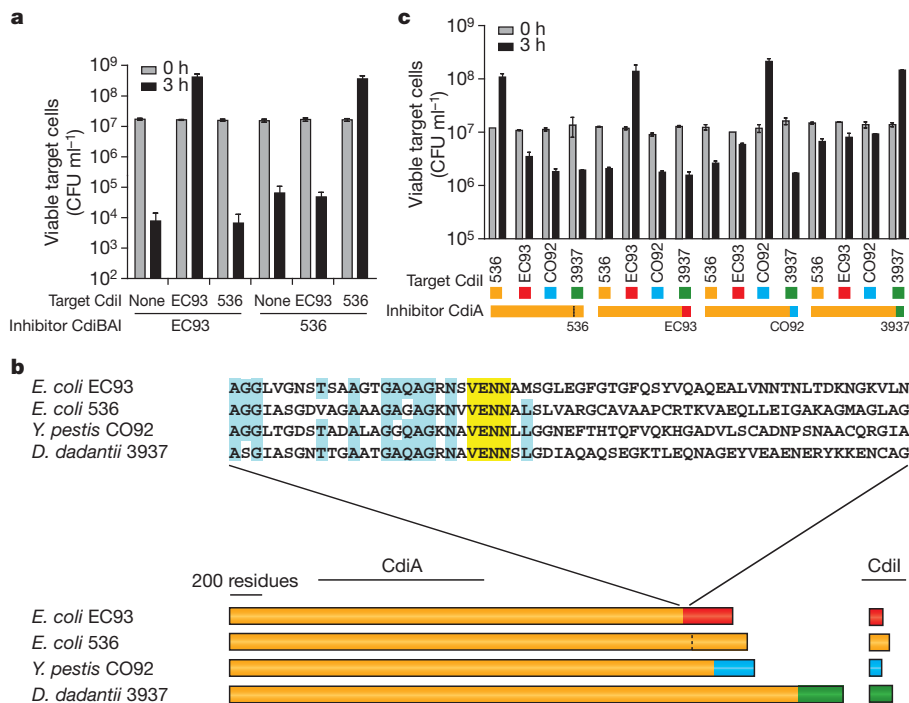
*E. coli* from growth inhibition mediated by cells expressing CdiA<sub>536</sub>, but not from cells expressing CdiA<sub>EC93</sub> (Fig. 1a). Similarly, CdiI<sub>EC93</sub> only provided immunity to cells expressing CdiA<sub>EC93</sub> (Fig. 1a). The protection conferred by CdiI therefore seems to be limited to its cognate CDI system. Alignment of CdiA<sub>EC93</sub> and CdiA<sub>536</sub> showed that about 3,000 residues at the amino terminus (up to and including a common VENN peptide motif) are 78% identical, but about 220 residues at the C terminus share no significant similarity<sup>6</sup> (Fig. 1b). To determine whether the dissimilar C termini of CdiA<sub>EC93</sub> and CdiA<sub>536</sub> account for the specificity of CdiI-mediated immunity, we replaced the coding sequences for CdiA-CT<sub>536</sub> and CdiI<sub>536</sub> in UPEC 536 with the corresponding region from EC93. The resulting strain produced a chimaeric CdiA protein in which the C-terminal 223 residues of CdiA<sub>EC93</sub> were fused to the N-terminal 3,020 residues of CdiA<sub>536</sub>. UPEC 536 producing this chimaeric CdiA inhibited target cells expressing CdiI<sub>536</sub> but not cells expressing CdiI<sub>EC93</sub>, whereas the converse was true for wild-type UPEC 536 (Fig. 1c). These results show that CdiA-CT<sub>EC93</sub> is functional when grafted to the CdiA molecule from UPEC 536, and that the CdiA-CT sequence is important for specificity of immunity.

The observation that CdiI-mediated immunity is specific to the CdiA-CT suggests that growth inhibitory activity is also contained within the CdiA-CT. Expression of the C-terminal 268 residues of CdiA<sub>EC93</sub> inside *E. coli* cells inhibited growth, and this inhibition was blocked by the co-production of CdiI<sub>EC93</sub> but not by that of CdiI<sub>536</sub> (Fig. 2a). Because the CdiA-CT lacks a secretion signal sequence, it is likely that this CdiA-CT-mediated growth inhibition and CdiI-mediated immunity occur within the cytoplasm. The minimal active region of CdiA-CT<sub>EC93</sub> was determined by deletion analysis. Removal of up to 25 residues from the N terminus of the 268-residue CdiA-CT<sub>EC93</sub> construct did not significantly affect growth inhibitory activity (Fig. 2b). Removal of 45 residues from the N terminus, which includes the conserved VENN motif, yielded a polypeptide with about tenfold greater inhibitory activity. Deletion of an additional 13 residues from the N terminus completely abrogated activity (Fig. 2b). Deletion of as few as 12 residues from the C terminus of the CdiA-CT<sub>EC93</sub> construct also abolished inhibitory activity. Thus, the growth inhibitory activity resides within the C-terminal 223 amino-acid residues of CdiA<sub>EC93</sub>.

We next searched for additional *cdiBAI* loci in other *E. coli* strains. Although several *cdiBAI* gene clusters are present in partly assembled *E. coli* genome sequences (Supplementary Table 1), we limited our analysis to the 33 fully assembled *E. coli* genomes currently available. Two *E. coli* isolates, UTI89 and CFT073, encode two-partner secretion systems related to EC93 CdiB/CdiA. The UTI89 CDI module is identical to that of UPEC 536 and is located within the same pathogenicity island (PAI<sub>II536</sub>) in both strains (Supplementary Fig. 2a)<sup>7</sup>. In contrast, the CFT073 CDI locus resides in an unrelated pathogenicity island (PAI-CFT073-*aspV*)<sup>5</sup>, and its predicted CdiA-CT and CdiI sequences show

<sup>1</sup>Department of Molecular, Cellular, and Developmental Biology, University of California – Santa Barbara (UCSB), Santa Barbara, California 93106-9625, USA. <sup>2</sup>Biomolecular Science and Engineering Program, University of California – Santa Barbara (UCSB), Santa Barbara, California 93106-9625, USA. <sup>†</sup>Present address: Department of Microbiology and Immunology, School of Medicine, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina 27599-7290, USA.

\*These authors contributed equally to this work.



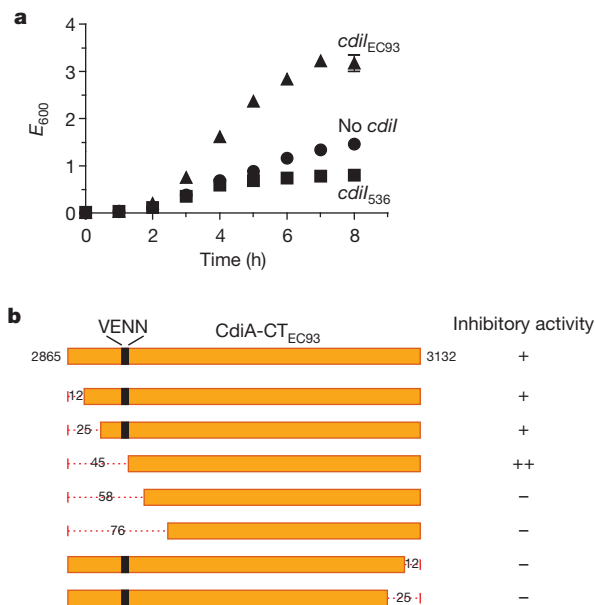
**Figure 1 | Analysis of CdiA chimaeras.** **a**, Target cells expressing CdiI<sub>EC93</sub> or CdiI<sub>536</sub> were co-cultured with inhibitor cells expressing *cdi* genes from either *E. coli* EC93 or UPEC 536. After 3 h, the number of viable target cells was determined as colony-forming units (CFU) per ml. Results are shown as means and s.d. ( $n = 4$  experiments). **b**, About 200 residues at the C terminus of the indicated CdiAs diverge after a conserved VENN peptide motif (shown as different colours). The CdiI proteins from each system are also highly variable. **c**, Target cells expressing CdiI from *E. coli* UPEC 536, *E. coli* EC93, *Y. pestis* CO92 (Uniprot accession no. Q7CGD9) and *D. dadantii* 3937 (CDI module 2; see Supplementary Table 1) were protected from CDI mediated by chimaeric CdiA<sub>536</sub> proteins containing cognate, but not heterologous, CdiA-CT. Results are shown as means and s.d. ( $n = 2$  experiments).

no similarity to the UPEC 536 or EC93 systems (Supplementary Figs 2b, c and 3b). The CdiA-CT of EC93 is unrelated to the CdiA-CTs identified in other fully sequenced *E. coli* strains; however, it is 42% identical to the CdiA-CT from another species, *Edwardsiella tarda* EIB202. This, together with the sporadic occurrence of *cdi* loci in *E. coli* strains and their association with genomic islands, suggests that these genes may be transferred horizontally.

Bioinformatic analyses showed that CDI systems are widespread among other Gram-negative bacteria, with representatives identified

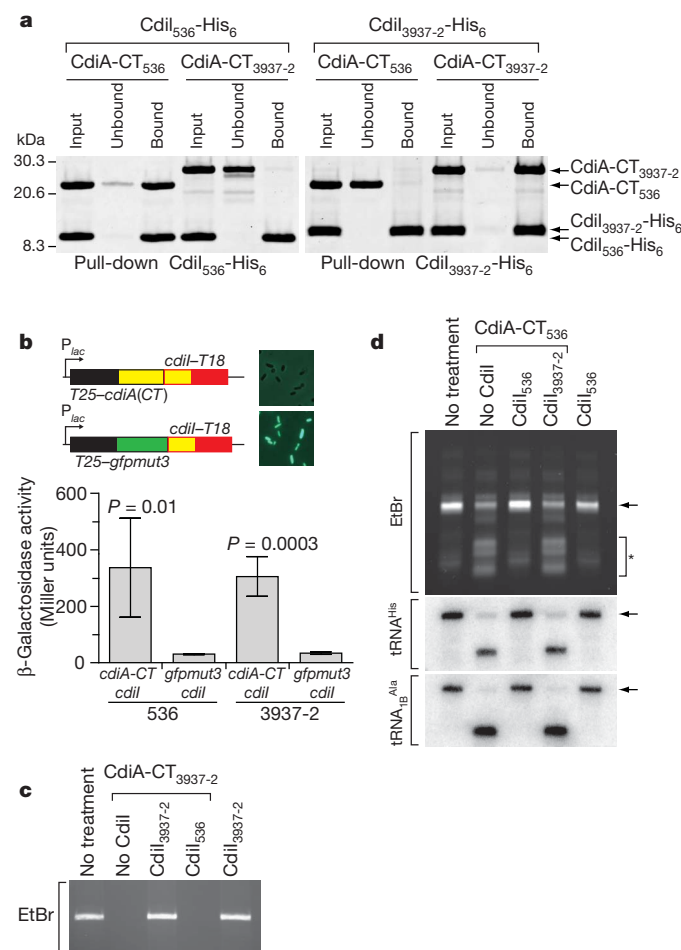
in a variety of  $\alpha$ -proteobacteria,  $\beta$ -proteobacteria and  $\gamma$ -proteobacteria (Supplementary Table 1 and Supplementary Fig. 3). Although widely distributed, only certain strains of any given species contain *cdiBAI* homologues. Some bacterial isolates encode multiple CDI modules; for example, *Bartonella grahamii* strain as4aup and *Photorhabdus luminescens* subsp. *laumondii* contain four and five CDI modules, respectively (Supplementary Fig. 3b). On the basis of their similarity to CdiA<sub>EC93</sub>, each putative *cdiA* homologue is likely to encode a TpsA family member with an N-terminal haemagglutination activity domain (Pfam PF05860, also called a TPS domain) and haemagglutinin repeats that are predicted to form a  $\beta$ -helical structure. In addition, most *cdiA* homologues encode the VENN peptide motif, which is part of the DUF638 domain of unknown function (Pfam PF04829) (Fig. 1b and Supplementary Fig. 3a). In general, significant variability between CdiA-CT/CdiI protein sequences was observed between different species and between different CdiA-CTs encoded within a single strain. There are instances in which an extended genomic region, including the CDI module, is conserved between different strains of the same species (as in *E. coli* UPEC 536 and UTI89, and most *Yersinia pestis* strains); in such cases the entire CdiA protein and CdiI are highly conserved. *Burkholderia* species also seem to encode CdiB/CdiA two-partner secretion systems, but these loci have a different gene organization (*cdiAIB* rather than *cdiBAI*) and the putative CdiA proteins lack the DUF638 domain. Instead of the VENN motif, the *Burkholderia* *cdiA* homologues encode an NxxLYN motif that precedes variable C-terminal domains (Supplementary Fig. 3c). In all instances, the *Burkholderia* *cdiA* genes are followed by short open reading frames, which may be analogous to the *cdiI* genes in *E. coli*.

To determine whether the *in silico*-identified *cdi* loci encode functional CDI systems, we replaced the CdiA-CT and CdiI coding regions of UPEC 536 with the corresponding sequences from *Y. pestis* CO92 and the region 2 CDI module from *D. dadantii* 3937 (Supplementary Table 1). UPEC 536 producing chimaeric CdiAs inhibited target cells expressing heterologous CdiI proteins but not cells expressing cognate CdiI (Fig. 1c), strongly suggesting that *Y. pestis* CO92 and *D. dadantii* 3937 encode CDI systems with allele-specific immunity proteins. These data also indicate that the N-terminal 3,020 residues of CdiA<sub>536</sub> are capable of delivering functional CdiA-CT domains from *Y. pestis* CO92 and *D. dadantii* 3937 into target cells.



**Figure 2 | CdiA-CT contains growth inhibitory activity.** **a**, Growth of *E. coli* cells (measured as attenuation at 600 nm ( $E_{600}$ )) expressing CdiA-CT<sub>EC93</sub> from plasmid pDAL778. Co-expression of cognate CdiI<sub>EC93</sub>, but not heterologous CdiI<sub>536</sub>, protected cells from growth inhibition. Results are shown as means  $\pm$  s.d. ( $n = 2$  experiments). **b**, The 268-residue CdiA-CT<sub>EC93</sub> peptide is depicted along with various truncation constructs indicating the number of residues deleted. Each CdiA-CT construct was tested for growth inhibitory activity when expressed in *E. coli* cells (+ +, growth was blocked immediately after CdiA-CT<sub>EC93</sub> induction; +, growth was blocked after a delay of 2–3 h; –, no growth inhibition).

How do CdiI immunity proteins protect against cognate CdiA-CT? We speculated that CdiI prevents CDI-mediated autoinhibition by binding specifically to the C terminus of cognate CdiA. To test this hypothesis, we examined the interaction between CdiA-CT and hexahistidine-tagged CdiI (CdiI-His<sub>6</sub>) proteins using Ni<sup>2+</sup>-affinity pull-down experiments. Because the CdiA-CTs used in these experiments lack His<sub>6</sub>-epitope tags, their retention on Ni<sup>2+</sup>-nitrilotriacetic acid (Ni<sup>2+</sup>-NTA) resin is dependent on binding to CdiI-His<sub>6</sub>. CdiA-CT<sub>536</sub> was retained by the Ni<sup>2+</sup>-NTA resin when preincubated with CdiI<sub>536</sub>-His<sub>6</sub> but not when preincubated with non-cognate CdiI<sub>3937-2</sub>-His<sub>6</sub> (Fig. 3a). Reciprocally, CdiA-CT<sub>3937-2</sub> bound to the resin only in the presence of cognate CdiI<sub>3937-2</sub>-His<sub>6</sub> (Fig. 3a). These data indicate



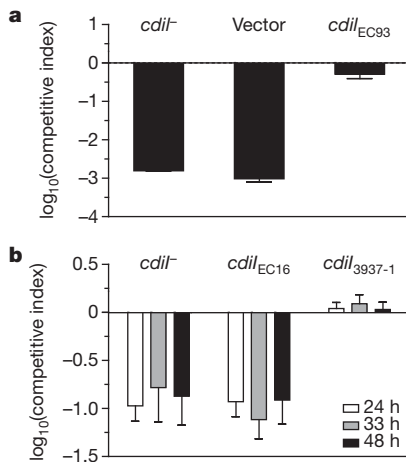
**Figure 3 | CdiI immunity protein binds specifically to cognate CdiA-CT and blocks activity.** **a**, Purified CdiA-CT and CdiI-His<sub>6</sub> proteins from UPEC 536 and *D. dadantii* 3937 (CDI module 2; 3937-2) were mixed *in vitro* with Ni<sup>2+</sup>-NTA resin. Aliquots of resin-bound and unbound fractions were analysed by SDS-PAGE and Coomassie blue staining. **b**, A bacterial two-hybrid system (BACTH) based on adenylate cyclase activity was used to monitor CdiA-CT/CdiI binding *in vivo*. A β-galactosidase reporter was used to measure adenylate cyclase activity. Expression of two-hybrid T25-cdiA-CT/cdiI-T18 fusion constructs resulted in significant β-galactosidase activity. T25-gfpmut3/cdiI-T18 fusions were used to control for background β-galactosidase activity. Fluorescence microscopy of cells expressing each *D. dadantii* 3937 construct confirms GFP expression of the control. *P* values were obtained with an unpaired, two-tailed *t*-test; results are shown as means ± s.d. (*n* = 2 experiments). **c**, Purified CdiA-CT<sub>3937-2</sub> was incubated with linear pUC19 DNA in the presence and absence of cognate and heterologous CdiI. Reactions were analysed by native agarose-gel electrophoresis and ethidium bromide (EtBr) staining. **d**, Purified CdiA-CT<sub>536</sub> was incubated with an S100 cell extract (100,000g supernatant), in the presence and absence of cognate and heterologous CdiI. Reactions were analysed by denaturing gel electrophoresis. Top panel: EtBr staining for total tRNA (arrow indicates tRNA; asterisk indicates degradation products). Lower panels, northern blot analyses of tRNA<sup>His</sup> and tRNA<sup>Ala</sup>. Arrows indicate full-length tRNAs.

that CdiI proteins bind to CdiA-CT *in vitro* in an allele-specific manner. To determine whether CdiA-CT and CdiI bind one another *in vivo*, we used a modified bacterial two-hybrid system. In this system, proteins of interest are fused to the T18 and T25 domains of adenylate cyclase, and binding of the two fusion proteins results in the production of cyclic AMP, which is monitored indirectly through the expression of a β-galactosidase reporter<sup>8</sup>. Co-expression of cognate T25-CdiA-CT<sub>536</sub> and CdiI<sub>536</sub>-T18 fusions, or T25-CdiA-CT<sub>3937-2</sub> and CdiI<sub>3937-2</sub>-T18 fusions, yielded high activities of β-galactosidase (more than 300 Miller units) in both cases (Fig. 3b). Because CdiA-CTs are cytotoxic in the absence of cognate CdiI, we used green fluorescent protein (GFP) as a negative control to test binding specificity. Co-expression of T25-GFP and CdiI<sub>536</sub>-T18, or T25-GFP and CdiI<sub>3937-2</sub>-T18, resulted in very low β-galactosidase activities (Fig. 3b). Taken together, these data demonstrate that CdiI immunity proteins bind to their cognate CdiA-CTs *in vitro* and *in vivo*.

How do CDI systems inhibit target cell growth? Most CdiA-CTs are not similar to known proteins or protein domains. However, we found that the C-terminal 132 residues of CdiA<sub>3937-2</sub> from *D. dadantii* 3937 shares 35% identity with the pyocin S3 nuclease domain from *Pseudomonas aeruginosa* (Supplementary Fig. 4). Pyocin S3 is cytotoxic by virtue of its DNase activity<sup>9</sup>, suggesting that CdiA<sub>3937-2</sub> may also use nuclease activity to inhibit target cell growth. We confirmed that purified CdiA-CT<sub>3937-2</sub> possesses a robust Mg<sup>2+</sup>-dependent DNase activity, capable of completely digesting linear and supercoiled plasmid DNA (Fig. 3c, and data not shown). We also examined the activity of CdiA-CT<sub>536</sub>, which does not share sequence homology with other known toxins or colicins, and found that it could cleave transfer RNA (Fig. 3d). Purified CdiA-CT<sub>536</sub> readily cleaved several *E. coli* tRNA species, but not ribosomal RNA or messenger RNA (Fig. 3d, and data not shown). For each CdiA-CT, the addition of purified cognate CdiI blocked nuclease activity, whereas the addition of heterologous CdiI had no effect on activity (Fig. 3c, d). These results suggest that CDI systems use more than one mechanism to inhibit cell growth. If these DNase and tRNase activities are responsible for growth inhibition, then the CdiA-CTs must be translocated into the target cell cytoplasm (Supplementary Fig. 1). According to this model, CdiI proteins confer immunity to CDI by binding to cognate CdiA-CTs and blocking their enzymatic activities. Given the diversity of CdiA-CT sequences (Fig. 1b and Supplementary Fig. 3), it seems likely that additional growth inhibitory mechanisms will be identified for other CDI systems.

It is not known when or where CDI systems are deployed in the environment, nor what precise biological function or functions they provide. UPEC 536 does not express the *cdiBAI* gene cluster under standard laboratory growth conditions (S.K.A., J.S.W. and D.A.L., unpublished observations). However, EC93 expresses *cdiBAI* constitutively<sup>2</sup>. To determine whether the CDI system in EC93 provides a selective advantage, we deleted the CdiA-CT and CdiI coding sequences of EC93 and mixed the resulting mutant cells with wild-type EC93 at a 1:1 ratio in a growth competition experiment. After 3 h of co-culture, EC93Δ*cdiA-CT*Δ*cdiI* mutant bacteria were less than 1% of wild-type EC93 cells (competitive index less than 10<sup>-2</sup>; Fig. 4a). However, EC93Δ*cdiA-CT*Δ*cdiI* cells expressing CdiI<sub>EC93</sub> from a plasmid were able to compete equally with wild-type EC93, indicating that the original loss of fitness was due to the mutant's susceptibility to CDI (Fig. 4a). These results indicate that CDI systems may be significant in intraspecies competition between bacteria occupying the same ecological niche. Further support for this conclusion came from an analysis of *D. dadantii*. Previous work has shown that disruption of a putative *cdiI* gene, designated *virA*, decreases the virulence of *D. dadantii* EC16 on plant hosts<sup>10</sup>. Our results indicate that VirA binds and inactivates the C-terminal domain of HecA<sub>EC16</sub>, a CdiA homologue (Supplementary Table 1, and data not shown). These results suggest that *D. dadantii* may express *cdi* genes on plants. We used the fully sequenced *D. dadantii* 3937 strain, which contains two CDI regions (*cdi*<sub>3937-1</sub> and *cdi*<sub>3937-2</sub>; see Supplementary Table 1), to test the hypothesis that CDI has a function





**Figure 4 | CDI systems function in intrastrain growth competition.**

**a**, Streptomycin-resistant (*str*<sup>R</sup>) *CDI*<sup>+</sup> EC93 cells were mixed 1:1 with rifampicin-resistant (*rif*<sup>R</sup>) EC93  $\Delta$ *cdiA-CT* $\Delta$ *cdiI* cells that either contained no plasmid (*cdiI*<sup>-</sup>), pBR322 plasmid (vector) or *CdiI*<sub>EC93</sub>-expressing plasmid (*cdiI*<sub>EC93</sub>). After 3 h of co-culture, cells were plated on Luria-Bertani medium, and *rif*<sup>R</sup> and *str*<sup>R</sup> were quantified (CFU ml<sup>-1</sup>) to calculate the competitive index [(*rif*<sup>R</sup> CFU/*str*<sup>R</sup> CFU)<sub>3h</sub>/(*rif*<sup>R</sup> CFU/*str*<sup>R</sup> CFU)<sub>0h</sub>]. Results are shown as means and s.d. (*n* = 2 experiments). *P* = 0.002. **b**, Gentamicin-resistant (*gent*<sup>R</sup>) *D. dadantii* 3937 cells were mixed 100:1 with nalidixic-acid-resistant (*nal*<sup>R</sup>) *D. dadantii* 3937  $\Delta$ *cdiA-CT*<sub>3937-1</sub>  $\Delta$ *cdiI*<sub>3937-1</sub> alone (*cdiI*<sup>-</sup>) or complemented with heterologous *cdiI* from *D. dadantii* EC16 (*cdiI*<sub>EC16</sub>) or cognate *cdiI* (*cdiI*<sub>3937-1</sub>). Cell mixtures were inoculated on chicory leaves (see Supplementary Methods) and incubated for the indicated durations, and viable counts were determined. The competitive index (*nal*<sup>R</sup> CFU/*gent*<sup>R</sup> CFU) was calculated as described for **a** at each time point. Results are shown as means and s.d. (*n* = 2 experiments). *P* value at 24 h = 0.00004. *P* values were obtained with an unpaired, two-tailed *t*-test.

in intrastrain competition. Each of the *cdiA-CT/cdiI* regions was deleted individually, and the resulting mutants were competed against wild-type *D. dadantii* 3937 on chicory<sup>11,12</sup>. Although deletion of the *cdiI*<sub>3937-2</sub> region had no effect on competition, cells lacking *cdiI*<sub>3937-1</sub> were outcompeted by wild-type bacteria, as demonstrated by a competitive index of about 10<sup>-1</sup> (Fig. 4b, and data not shown). This competitive disadvantage was reversed by complementation with a chromosomal copy of the cognate *cdiI*<sub>3937-1</sub> gene (Fig. 4b). Complementation was specific, because non-cognate *cdiI*<sub>EC16</sub> (*virA*) from *D. dadantii* EC16 had no effect on the competitive index (Fig. 4b). These results show that the region 1 CDI system in *D. dadantii* has a function in growth competition on chicory. The role of the region 2 CDI system is unknown, but it could function under different environmental conditions or target different bacterial species. Taken together, these results strongly indicate that CDI systems function in growth competition in the environment.

## METHODS SUMMARY

Strains, plasmids and oligonucleotides used in this study are shown in Supplementary Table 2. *E. coli* competition assays were performed as described previously<sup>2</sup>. *D. dadantii* competition assays on chicory were performed as described in Supplementary Methods. EC93 and *D. dadantii* 3937 *cdiA-CT-cdiI* deletions and *CdiA* chimaeras were constructed with allelic exchange as described previously<sup>13</sup>. For chimaera construction, the 3' end of *cdiA* and all of *cdiI* from UPEC 536  $\Delta$ *kpsK15*  $\Delta$ *araCBA* *specRExBAD-cdiBAI* (DL5646) were replaced with *cdiA-CT* (sequence immediately following VENNX) and *cdiI* from *E. coli* EC93, *Y. pestis* CO92 (Uniprot accession number Q7CGD9), or *D. dadantii* 3937 region 2 (see Supplementary Methods). The  $\Delta$ *kpsK15* capsule mutation was used to increase the efficacy of CDI, on the basis of our previous results showing that capsule production blocks CDI<sup>3</sup>. Immunity plasmids were constructed by ligating PCR-amplified *cdiI* genes into plasmid pBR322 under *tet* promoter control (Fig. 1a, c). For *D. dadantii* 3937 the immunity plasmids were constructed by ligating PCR-amplified *cdiI* genes into the miniTn7 delivery plasmid pUC18R6KT-miniTn7T under *tet* promoter control (see Supplementary Methods). Deletion mapping of *E. coli* EC93 *cdiA-CT* (Fig. 2b) was performed by cloning specific sequences amplified by PCR into

plasmid pLAC11 (ref. 14) under *lac* promoter control. All plasmids were propagated in EPI100 *acrB* mutant strain DL5154 to mitigate toxic effects. *In vivo* interactions between *CdiA-CT* and *CdiI* were determined with a modified bacterial two-hybrid system (BACTH; Euromedex)<sup>8</sup>.  $\beta$ -Galactosidase<sup>14</sup> and fluorescence<sup>3</sup> analyses were performed as described previously. *In vitro* affinity pull-downs with His<sub>6</sub>-tagged *CdiI/CdiA-CT* were performed with Ni<sup>2+</sup>-NTA resin (Qiagen) (Fig. 3a). *CdiA-CT* was released by denaturation in buffer containing 6 M guanidinium-HCl, and His<sub>6</sub>-tagged *CdiI* was released in native buffer supplemented with 250 mM imidazole. *CdiA-CT* activities were analysed as described in Supplementary Methods. Complete methods are presented in Supplementary Methods.

Received 21 January; accepted 6 September 2010.

- Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nature Rev. Microbiol.* **8**, 15–25 (2010).
- Aoki, S. K. *et al.* Contact-dependent inhibition of growth in *Escherichia coli*. *Science* **309**, 1245–1248 (2005).
- Aoki, S. K. *et al.* Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Mol. Microbiol.* **70**, 323–340 (2008).
- Aoki, S. K., Webb, J. S., Braaten, B. A. & Low, D. A. Contact-dependent growth inhibition causes reversible metabolic downregulation in *Escherichia coli*. *J. Bacteriol.* **191**, 1777–1786 (2009).
- Luo, C., Hu, G. Q. & Zhu, H. Genome reannotation of *Escherichia coli* CFT073 with new insights into virulence. *BMC Genomics* **10**, 552–561 (2009).
- Pagni, M. & Jongeneel, C. V. Making sense of score statistics for sequence alignments. *Brief. Bioinform.* **2**, 51–67 (2001).
- Dobrindt, U. *et al.* Genetic structure and distribution of four pathogenicity islands (PAI<sub>1536</sub> to PAI<sub>IV536</sub>) of uropathogenic *Escherichia coli* strain 536. *Infect. Immun.* **70**, 6365–6372 (2002).
- Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl Acad. Sci. USA* **95**, 5752–5756 (1998).
- Duport, C., Baysse, C. & Michel-Briand, Y. Molecular characterization of pyocin S3, a novel S-type pyocin from *Pseudomonas aeruginosa*. *J. Biol. Chem.* **270**, 8920–8927 (1995).
- Rojas, C. M. *et al.* The *Erwinia chrysanthemi* EC16 *hrp/hrc* gene cluster encodes an active Hrp type III secretion system that is flanked by virulence genes functionally unrelated to the Hrp system. *Mol. Plant Microbe Interact.* **17**, 644–653 (2004).
- Ham, J. H. *et al.* Analysis of *Erwinia chrysanthemi* EC16 *pelE:uidA*, *pelL:uidA*, and *hrpN:uidA* mutants reveals strain-specific atypical regulation of the Hrp type III secretion system. *Mol. Plant Microbe Interact.* **17**, 184–194 (2004).
- Yap, M. N., Rojas, C. M., Yang, C. H. & Charkowski, A. O. Harpin mediates cell aggregation in *Erwinia chrysanthemi* 3937. *J. Bacteriol.* **188**, 2280–2284 (2006).
- Edwards, R. A., Keller, L. H. & Schifferli, D. M. Improved allelic exchange vectors and their use to analyze 987P fimbria gene expression. *Gene* **207**, 149–157 (1998).
- Braaten, B. A., Nou, X., Kaltenbach, L. S. & Low, D. A. Methylation patterns in *pap* regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. *Cell* **76**, 577–588 (1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Charkowski, A. Collmer, J. Roth and H. Schweizer for plasmids, bacterial strains and helpful discussions; W. Latham for *Y. pestis* CO92 DNA; and R. Christoffersen for helpful advice on plant experiments. This work was supported by National Science Foundation grant 0642052 (D.A.L.), a Tri-Counties Blood Bank Postdoctoral Fellowship (S.K.A.) and National Institutes of Health grants GM078634 (C.S.H.), AI043986 (P.A.C.) and U54AI065359 (D.A.L., P.A.C. and C.S.H.). The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases or the National Institutes of Health. This project made use of preliminary sequences from the *Dickeya dadantii* 3937 genome project supported by the Initiative for Future Agriculture and Food Systems grant no. 2001-52100-11316 from the United States Department of Agriculture Cooperative State Research, Education, and Extension Service.

**Author Contributions** D.A.L., S.K.A., P.A.C. and C.S.H. designed the research. D.A.L., C.S.H., P.A.C., S.J.P. and S.K.A. prepared the manuscript. S.J.P. and B.R.B. performed bioinformatic analyses. B.R.B., A.M.J. and P.A.C. obtained initial evidence for the toxic nature of *CdiA-CT*, variability of *CdiA-CTs* and *CdiIs*, and binding between *CdiA-CTs* and cognate *CdiIs* in studies of *Burkholderia pseudomallei* *cdi* genes. S.K.A. cloned and performed competition assays and growth curves involving *cdiBAI*, *cdiA-CT* and *cdiA* chimaeras and deletions. B.A.B. and J.S.W. conducted the deletion mapping study, with assistance from S.K.A. B.A.B. and S.K.A. conducted the bacterial two-hybrid study. C.T.R. constructed *D. dadantii* *cdi* mutants and plasmids and performed growth competition assays on chicory. E.J.D. cloned, purified protein and performed the *in vitro* protein interaction and *CdiA-CT* activity studies.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.A.L. (low@lifesci.ucsf.edu).

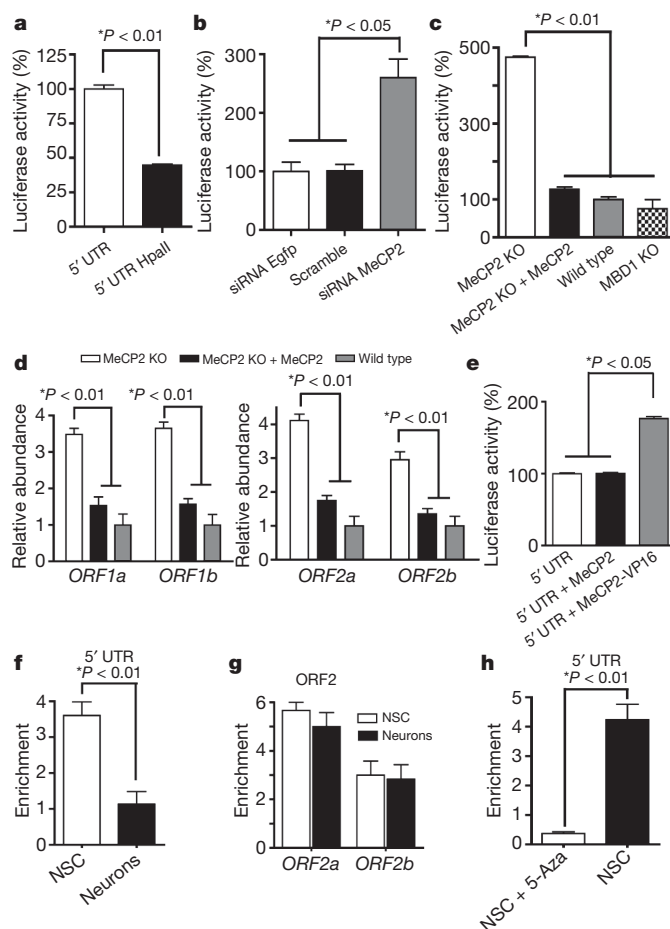
# L1 retrotransposition in neurons is modulated by MeCP2

Alysson R. Muotri<sup>1\*</sup>, Maria C. N. Marchetto<sup>2\*</sup>, Nicole G. Coufal<sup>2</sup>, Ruth Oefner<sup>2</sup>, Gene Yeo<sup>3</sup>, Kinichi Nakashima<sup>4</sup> & Fred H. Gage<sup>2</sup>

Long interspersed nuclear elements-1 (LINE-1 or L1s) are abundant retrotransposons that comprise approximately 20% of mammalian genomes<sup>1–3</sup>. Active L1 retrotransposons can impact the genome in a variety of ways, creating insertions, deletions, new splice sites or gene expression fine-tuning<sup>4–6</sup>. We have shown previously that L1 retrotransposons are capable of mobilization in neuronal progenitor cells from rodents and humans and evidence of massive L1 insertions was observed in adult brain tissues but not in other somatic tissues<sup>7,8</sup>. In addition, L1 mobility in the adult hippocampus can be influenced by the environment<sup>9</sup>. The neuronal specificity of somatic L1 retrotransposition in neural progenitors is partially due to the transition of a Sox2/HDAC1 repressor complex to a Wnt-mediated T-cell factor/lymphoid enhancer factor (TCF/LEF) transcriptional activator<sup>7,10</sup>. The transcriptional switch accompanies chromatin remodelling during neuronal differentiation, allowing a transient stimulation of L1 transcription<sup>7</sup>. The activity of L1 retrotransposons during brain development can have an impact on gene expression and neuronal function, thereby increasing brain-specific genetic mosaicism<sup>11,12</sup>. Further understanding of the molecular mechanisms that regulate L1 expression should provide new insights into the role of L1 retrotransposition during brain development. Here we show that L1 neuronal transcription and retrotransposition in rodents are increased in the absence of methyl-CpG-binding protein 2 (MeCP2), a protein involved in global DNA methylation and human neurodevelopmental diseases. Using neuronal progenitor cells derived from human induced pluripotent stem cells and human tissues, we revealed that patients with Rett syndrome (RTT), carrying MeCP2 mutations, have increased susceptibility for L1 retrotransposition. Our data demonstrate that L1 retrotransposition can be controlled in a tissue-specific manner and that disease-related genetic mutations can influence the frequency of neuronal L1 retrotransposition. Our findings add a new level of complexity to the molecular events that can lead to neurological disorders.

In neural stem cells, the repressor complex on the L1 promoter region (L1 5'UTR) includes the transcriptional factor Sox2 and the histone deacetylase 1 protein (HDAC1)<sup>7</sup>, a MeCP2 partner<sup>13,14</sup>. MeCP2 has been shown to interfere with the L1 5'UTR promoter activity in transformed cell lines<sup>15</sup>. To investigate the role of MeCP2 in the activity of L1 promoter in neural stem cells, we cloned the L1 promoter region upstream to the luciferase gene, generating the L1 5'UTR–Luc plasmid<sup>7</sup>. Methylation of the L1 5'UTR–Luc reduced the promoter activity in neural stem cells (Fig. 1a and Supplementary Fig. 1a). Reduction of MeCP2 levels using siRNAs led to an increase in luciferase activity (Fig. 1b and Supplementary Fig. 1b). Transfection of the L1 5'UTR–Luc methylated plasmid in mouse neuroepithelial cells revealed that the L1 promoter activity was approximately four times more active in the MeCP2 knockout (KO) background than in wild-type (Fig. 1c and Supplementary Fig. 1c). Ectopic MeCP2 expression reduced the luciferase activity in MeCP2 KO cells close to wild-type levels (Fig. 1c).

We repeated the luciferase assay using neuroepithelial cells from a sibling MBD1 KO animal<sup>16</sup>. MBD1 (methyl-CpG binding domain protein 1) is part of the methyl-binding protein family and has differential



**Figure 1 | MeCP2 silences L1 expression.** **a**, Methylation of the L1 5'UTR–Luc reduced its transcriptional activity. **b**, Reduction of MeCP2 transcripts correlates with increased L1 promoter activity. **c**, Increased L1 promoter activity in the absence of MeCP2 but not MBD1. **d**, L1 RNA levels correlate with MeCP2 expression. **e**, Expression of the MeCP2–VP16 increased the activity of the L1 5'UTR promoter. **f**, **g**, Recruitment of MeCP2 on L1 sequences by ChIP in neural stem cells (NSC) or neurons, using 5'UTR primers (**f**) and two ORF2 regions (**g**). **h**, Occupancy of MeCP2 on the L1 promoter requires DNA methylation. Removal of DNA methylation with 5-azacytidine (5-Aza) reduced MeCP2 association to L1 promoter. ChIP–qPCR shows enrichment over IgG control precipitation. All experiments show experimental triplicates. Error bars in all panels show s.e.m.

<sup>1</sup>University of California San Diego, School of Medicine, Department of Pediatrics/Rady Children's Hospital San Diego, Department of Cellular & Molecular Medicine, Stem Cell Program, 9500 Gilman Drive, La Jolla, California 92093-0695, USA. <sup>2</sup>Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>3</sup>University of California San Diego, School of Medicine, Department of Cellular & Molecular Medicine, Stem Cell Program, 9500 Gilman Dr, La Jolla, California 92093-0695, USA. <sup>4</sup>Laboratory of Molecular Neuroscience, Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0101, Japan.

\*These authors contributed equally to this work.

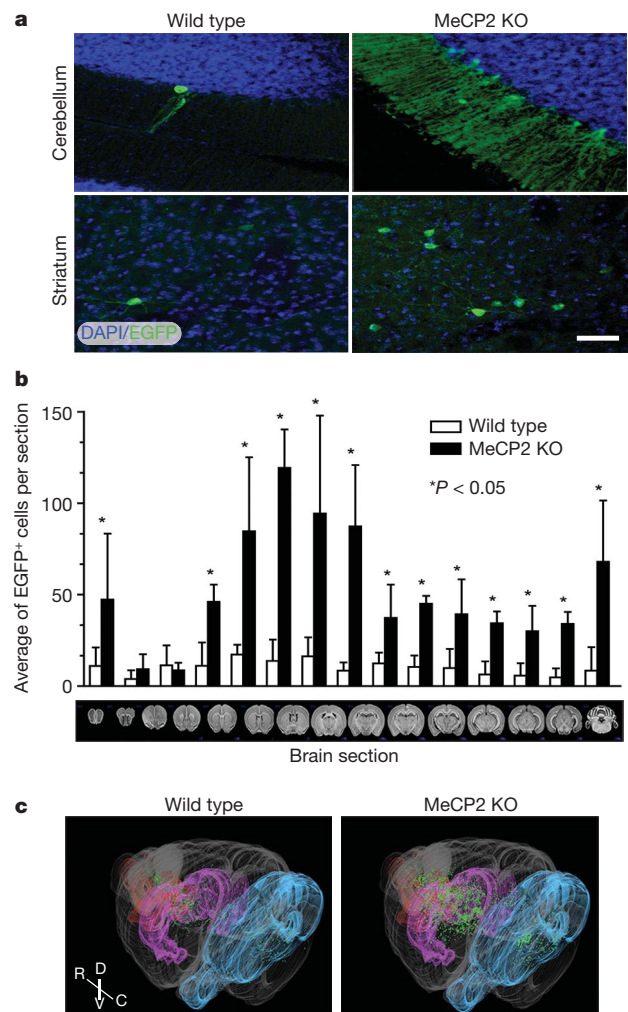


DNA specificity when compared to MeCP2<sup>17</sup>. The L1 promoter was not activated in MBD1 KO background, a finding that is consistent with the idea that L1 transcriptional repression is specific to MeCP2 (Fig. 1c). Moreover, the promoter activity correlated well with the level of L1 RNA, as measured by qPCR (Fig. 1d). Ectopic MeCP2 expression reduced L1 RNA levels in the MeCP2 KO background (Fig. 1d). We co-transfected neural stem cells with the methylated L1 5'UTR-Luc and a plasmid containing either the MeCP2 cDNA or the MeCP2 fused with the transactivator domain VP16. The overexpression of MeCP2 alone did not change the luciferase levels, but the MeCP2-VP16 fusion increased luciferase levels twofold (Fig. 1e).

Using chromatin immunoprecipitation (ChIP) followed by quantitative PCR (qPCR), we detected high levels of MeCP2 in association with endogenous L1 promoter regions in neural stem cells compared to neurons (Fig. 1f). MeCP2 was also associated with other L1 regions (ORF2), but this association did not change during differentiation (Fig. 1g; see controls for ChIP experiments in Supplementary Fig. 1d, e). After treatment with 5-azacytidine, the MeCP2 ChIP signal was reduced and L1 expression increased (Fig. 1h and Supplementary Fig. 1f). A set of the CpG sites within the L1 promoter had a tendency to demethylate during neuronal differentiation, indicating that DNA methylation may silence L1 expression in neural stem cells by attracting MeCP2 (Supplementary Fig. 1g, h).

To study L1 regulation *in vivo*, we compared the brains of the L1-EGFP (enhanced green fluorescent protein) transgenic mice in wild-type and MeCP2 KO backgrounds. L1-EGFP transgenic mice have a L1 indicator cassette that will only activate the EGFP reporter after retrotransposition<sup>7</sup> (Supplementary Fig. 2a). The numbers of EGFP-positive cells in the brains of MeCP2 KO mice were significantly higher than in wild type (Fig. 2a, b). EGFP-positive cells were also observed in the germ line of MeCP2 KO at similar frequency as in wild-type animals, but not in other somatic tissues (Supplementary Fig. 2b). To visualize the distribution of EGFP-positive cells, we generated high-resolution, three-dimensional maps of both MeCP2 KO and wild-type brains. Although MeCP2 KO brain sections had an average of 3.5-fold more EGFP-positive cells than wild type, certain brain structures were more prone to L1 retrotransposition (Fig. 2b, c). Specifically, the cerebellum, striatum, cortex, hippocampus and olfactory bulb contained 4.2-, 5.3-, 2.8-, 6.3- and 3.8-fold more EGFP-positive neurons, respectively, in the MeCP2 KO genetic background than in wild type (Supplementary Fig. 3 and Supplementary Movie). More EGFP-positive cells may suggest an increased rate of L1 retrotransposition and/or higher rate of MeCP2 KO cell proliferation with the newly retrotransposed EGFP reporter. We found no evidence that neuroepithelial cells from the MeCP2 KO genetic background had a higher rate of division than wild type (Supplementary Fig. 4a).

We next asked whether endogenous L1 retrotransposition was also increased in the MeCP2 KO brain. New insertions from retroelements can be quantified using a qPCR approach<sup>8,18</sup>. To determine the activity of endogenous L1 elements, we developed a technique based on single-cell genomic qPCR that measures the frequency of mouse L1 sequences within the genome (Fig. 3a). We proposed that MeCP2 KO-derived neuroepithelial cells would have increased genomic content of L1 sequences compared to wild-type cells. Neuroepithelial cells from wild-type and MeCP2 KO sibling mouse embryos were synchronized in G1 phase and karyotyped, to avoid interference during genomic L1 detection (Supplementary Fig. 4b, c). Finally, single-cell amplification using primers for ORF2 from active L1 families confirmed the presence of the expected amplicons (Supplementary Fig. 4d). MeCP2 KO-derived neuroepithelial cells displayed significantly more ORF2 genomic copies than wild-type cells (Fig. 3b). Specific primers for the L1 5'UTR were also tested in neuroepithelial cells and did not reveal an increase in copy number in MeCP2 KO background (Fig. 3c). This lack of difference can be explained by the fact that, upon retrotransposition, the 5' region of the L1 sequence is frequently truncated<sup>19,20</sup>. Also, no difference between genetic backgrounds was observed when using primers for non-mobile



**Figure 2 | MeCP2 modulates neuronal L1 retrotransposition *in vivo*.**

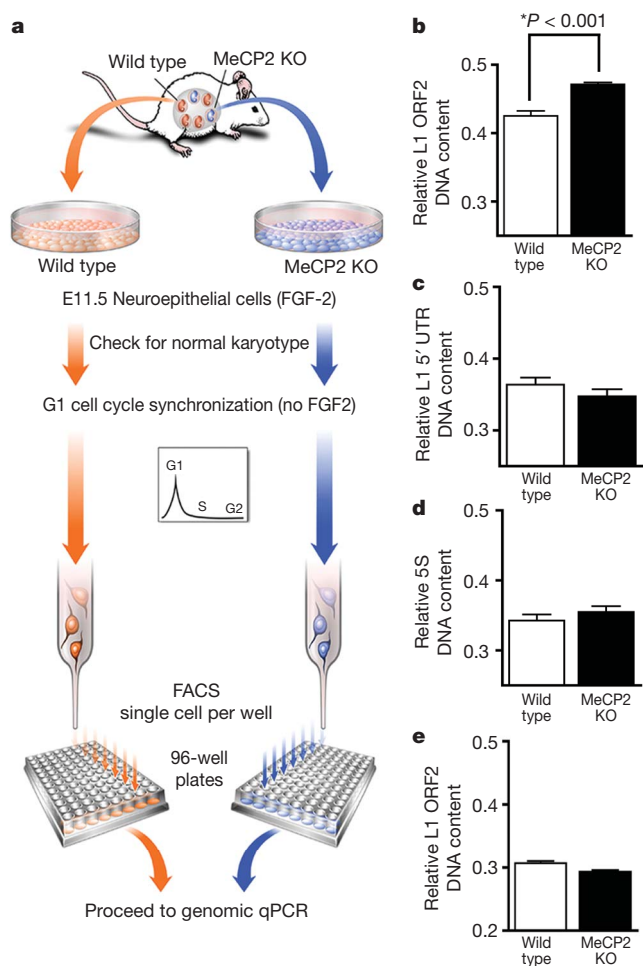
**a**, EGFP-positive cells, indicating *de novo* L1 retrotransposition, were found in several regions of the brain. The images were taken from sections that were highly affected by L1 retrotransposition. Bar, 30  $\mu$ m. **b**, Quantification of brain sections in MeCP2 KO background revealed more EGFP-positive cells compared to wild type ( $n = 6$  animals for each group). Error bars show s.d. **c**, Representative images from a three-dimensional reconstruction of wild-type and MeCP2 KO brains carrying the L1-EGFP transgene. Single dots (green) represent neurons that supported L1-EGFP retrotransposition. Olfactory bulb is shown in red, striatum in magenta and cerebellum in cyan. R, rostral; C, caudal; D, dorsal and V, ventral.

5S ribosomal RNA repetitive sequences (Fig. 3d). Another control experiment was performed using fibroblasts isolated from the two backgrounds (Fig. 3e). We did not observe a highly significant increase in L1 copy number in MeCP2 KO compared to wild type fibroblasts.

Mutations on the MeCP2 gene cause RTT, characterized by arrested development in early childhood and autistic behaviour at different levels of intensity<sup>21</sup>. To determine if L1 retrotransposition could occur in neuronal progenitor cells (NPC) derived from RTT patients, we generated induced pluripotent stem cells (iPSC) from a RTT patient's fibroblasts carrying a frameshift MeCP2 mutation and from a control, non-affected individual. All clones were pluripotent and able to produce NPC and neurons (Supplementary Fig. 5). Thus, we tested if the iPSC-derived NPC supported L1 retrotransposition.

NPC from both wild-type and RTT iPSC expressed the neural markers Sox1, Musashi1, Nestin and Sox2 at similar rates at the time of the experiment (Supplementary Fig. 6a, b). RTT and wild-type cells were electroporated with the L1<sub>RE3</sub>-EGFP reporter construct<sup>22,23</sup> (Fig. 4a). EGFP expression was detected in both wild-type and RTT cells (Fig. 4b). The frequency of EGFP-positive cells was approximately

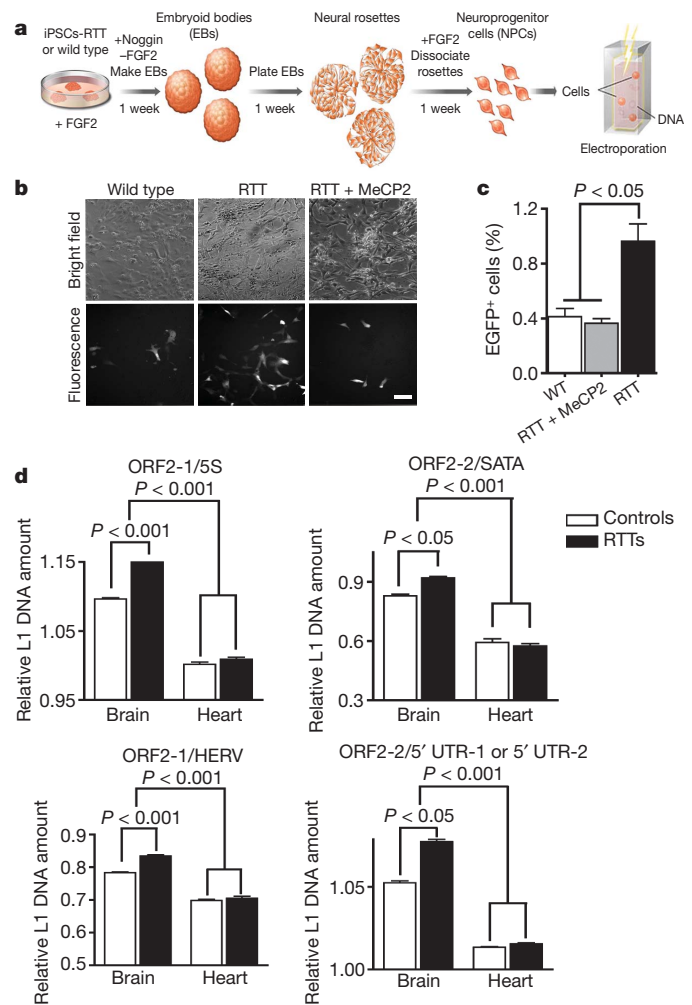




**Figure 3 | Endogenous L1 retrotransposition in mouse neuroepithelial cells.** **a**, Neuroepithelial cells harvested from embryonic day 11.5 (E11.5) sibling embryos were synchronized and sorted in individual wells followed by qPCR. **b**, Neuroepithelial cells in the MeCP2 KO background had higher L1 ORF2 DNA content than wild-type cells ( $P < 0.001$ ). **c**, L1 5' UTR primers did not reveal a significant increase in copy number in MeCP2 KO background. **d**, Non-mobile 5S ribosomal genes were used as controls. **e**, The difference in the amount of L1 ORF2 DNA in fibroblasts from the different genetic backgrounds was smaller than in the neural lineage. All experiments show experimental triplicates ( $n = 192$  cells for each primer pair). Error bars in all panels show s.e.m.

twofold higher in RTT than in control cells. Moreover, MeCP2 complementation reduced the levels of EGFP-positive cells in RTT NPC (Fig. 4b, c and Supplementary Fig. 6c). PCR confirmed the presence of the retrotransposed EGFP and sequencing confirmed the precise splicing of the intron (Supplementary Fig. 6d). We concluded that L1 activity could be facilitated by loss of MeCP2 function in human cells. We extended the iPSC findings *in vivo* using post mortem human tissues. To analyse the amounts of L1 retrotransposition in RTT patients and controls, brain and heart tissue was obtained from the same individuals. After genomic DNA extraction, a qPCR was used to compare the number of L1 ORF2 sequences normalized by four distinct non-mobile repetitive sequences. The number of L1 ORF2 sequences in the brains of RTT patients was significantly higher than in age/gender-matched controls (Fig. 4d). Moreover, the number of ORF2 sequences was higher in brain tissues in both controls and RTT patients when compared to heart tissue from the same individuals.

Our findings support previous data demonstrating that L1 5' UTR sequences are MeCP2 targets that may be subjected to methylation-dependent repression<sup>15,17</sup>. However, we cannot exclude an indirect effect of MeCP2 in regulating genes involved in L1 expression and/



**Figure 4 | L1 retrotransposition in RTT patients.** **a**, Schematic view of the NPC differentiation from iPSC followed by L1<sub>RE3</sub>-EGFP electroporation. **b**, Representative images of iPS-derived NPC expressing EGFP after L1 retrotransposition. Bar, 30  $\mu$ m. **c**, Quantification of the EGFP-positive cells after transfection. **d**, Primers for ORF2 were used to multiplex with primers for control sequences, such as the 5S ribosomal gene (5S), the satellite alpha (SATA) region, the human endogenous retrovirus H (HERV) sequence and the 5' UTR. The inverse ratio of ORF2/5S represents the amount of L1 ORF2 sequence in each sample ( $n = 5$  individuals per group). Similar results were obtained when different primers/probe for ORF2 (ORF2-2) were multiplex/normalized to other control sequences, using two pair of primers (5' UTR-1 or 5' UTR-2). Error bars show s.e.m., and the experiments were performed in triplicate.

or in changing the chromatin epigenetic landscape to facilitate *de novo* L1 insertions. An additive effect of multiple mechanisms is likely. Using different strategies, we have shown that L1 retrotransposition can be modulated by MeCP2. First, we demonstrated that MeCP2 can downregulate L1 promoter activity. Second, L1 retrotransposition from the L1-EGFP transgenic mice was significantly higher in the brains of a MeCP2 KO background than in a wild-type sibling animal. The L1-EGFP indicator system underestimates the actual capacity of retrotransposition and does not take into account insertions that truncate or silence the reporter cassette, *in trans* retrotransposition of *Alu* sequences or other RNAs<sup>24-26</sup>. Third, we developed a new technique based on single-cell genomic qPCR to measure the relative abundance of L1 sequences, revealing that MeCP2 KO neuroepithelial cells have more L1 sequences in the genome than wild-type cells. Lastly, RTT-NPC showed a higher L1 retrotransposition frequency than control cells. A qPCR experiment extended these observations to human brain samples from RTT patients compared to controls.

Our data provide evidence of a role for DNA methylation-dependent MeCP2 activity in controlling L1 mobility in the nervous system. Re-activation of MeCP2 expression was shown to reverse some of the neurological symptoms in MeCP2 KO mice<sup>27</sup>. The high rates of neuronal retrotransposition in the MeCP2 KO mice and RTT patients may be a consequence, rather than a cause, of the disease process. Nonetheless, new somatic insertions, especially at early developmental stages, may contribute to the genetic and epigenetic status of mature neurons at later stages of life. Early developmental structural and functional modulations could have potential consequences for RTT, where the detrimental effects of MeCP2 mutation occur at later postnatal stages. It is plausible to conclude that the RTT process leads to an increased rate of somatic mutations in the brain. Increased L1 neuronal retrotransposition is a novel and unexpected characteristic of RTT pathology. Our findings add a new layer of complexity to the understanding of genomic plasticity and may have direct implications for individual variation and for neurological diseases.

## METHODS SUMMARY

For the *luciferase* activity experiments, rat neural stem cells were isolated, characterized and cultured as described<sup>28</sup>. Neuroepithelial cells from time-pregnant midgestation (embryonic day 11.5) telencephalons from male wild-type, MBD1 KO, and MeCP2 KO sibling mouse embryos, from the same genetic background (C57BL/6J) were isolated. Cells were cultured for two to three passages in Dulbecco's modified Eagle's medium (DMEM) F12 media with N2 supplement and fibroblast growth factor 2 (FGF2) as described elsewhere<sup>29</sup>. Plasmid and siRNA transfections were performed by electroporation (Lonza/Amara Biosystem). Luciferase activity was measured with the Dual-Luciferase reporter assay system (Promega) according to the manufacturer's protocol. Chromatin immunoprecipitation (ChIP) assays were performed following the manufacturer's protocol using a kit from Millipore/Upstate. Antibodies used were anti-MeCP2 and IgG (Upstate). After immunoprecipitation, recovered chromatin fragments were subjected to PCR using primers for the rat L1 sequence. qPCR values were normalized to the IgG precipitation and shown as fold enrichment. For human iPSC derivation, RTT and control fibroblasts were infected with retroviral vectors containing the *Oct4*, *c-Myc*, *Klf4* and *Sox2* human cDNAs as described previously by Yamanaka's group<sup>30</sup>. iPSC-derived neural progenitors were electroporated (Lonza/Amara Biosystem) with L1-EGFP plasmid and FACS sorted for EGFP to quantify L1 *de novo* insertions. Single-cell genomic quantitative PCR (qPCR) was performed in cell-cycle-arrested neuroepithelial cells and fibroblasts from wild-type and MeCP2 KO mice. The plates containing one cell per well were then snap frozen at  $-80^{\circ}\text{C}$  until the day of the qPCR. The qPCR was performed using the protocol available on the manufacturer's website (Applied Biosystems). Briefly, a solution containing forward/reverse primers and SYBR Green PCR Master Mix was added to the previously sorted cells and the detection of DNA products was carried out in an ABI PRISM 7900HT Sequence Detection System. For multiplex genomic qPCR in human tissues the qPCR strategy and L1 copy estimation were done as previously described<sup>8</sup>.

Received 27 April; accepted 30 September 2010.

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Gibbs, R. A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
3. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
4. Kazazian, H. H. Jr. Mobile elements and disease. *Curr. Opin. Genet. Dev.* **8**, 343–350 (1998).
5. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274 (2004).
6. Perepelitsa-Belancio, V. & Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nature Genet.* **35**, 363–366 (2003).

7. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
8. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131 (2009).
9. Muotri, A. R., Zhao, C., Marchetto, M. C. & Gage, F. H. Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* **19**, 1002–1007 (2009).
10. Kuwabara, T. *et al.* Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nature Neurosci.* **12**, 1097–1105 (2009).
11. Muotri, A. R. & Gage, F. H. Generation of neuronal variability and complexity. *Nature* **441**, 1087–1093 (2006).
12. Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G. & Gage, F. H. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* **33**, 345–354 (2010).
13. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
14. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.* **19**, 187–191 (1998).
15. Yu, F., Zingler, N., Schumann, G. & Straling, W. H. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* **29**, 4493–4501 (2001).
16. Zhao, X. *et al.* Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc. Natl Acad. Sci. USA* **100**, 6777–6782 (2003).
17. Klose, R. J. *et al.* DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* **19**, 667–678 (2005).
18. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
19. Grimaldi, G., Skowronski, J. & Singer, M. F. Defining the beginning and end of *KpnI* family segments. *EMBO J.* **3**, 1753–1759 (1984).
20. Moran, J. V. & Gilbert, N. in *Mobile DNA II*, Vol. 2 (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) Ch. 35, 836–869 (ASM Press, 2002).
21. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23**, 185–188 (1999).
22. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
23. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28**, 1418–1423 (2000).
24. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**, 363–367 (2000).
25. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet.* **35**, 41–48 (2003).
26. Wei, W. *et al.* Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell Biol.* **21**, 1429–1439 (2001).
27. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315**, 1143–1147 (2007).
28. Palmer, T. D., Takahashi, J. & Gage, F. H. The adult rat hippocampus contains primordial neural stem cells. *Mol. Cell Neurosci.* **8**, 389–404 (1997).
29. Nakashima, K. *et al.* Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* **284**, 479–482 (1999).
30. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** A.R.M. is supported by the National Institutes of Health through the NIH Director's New Innovator Award Program, 1-DP2-OD006495-01 and by the Emerald Foundation. F.H.G. is supported by the Mathers Foundation, Lookout Fund, and NIH/NINDS R01MH088485. The authors would like to thank A. Huynh, B. Aimone, K. Stecker, B. Berg and D. Sepp for help during the 3D brain model assembly, J. Moran and J. Garcia-Perez for discussion and critical review of the manuscript, M. Gage for editorial comments, and B. Moddy and G. Peng for experimental assistance.

**Author Contributions** A.R.M. and M.C.N.M. are the leading authors. They contributed to the concept, designed and performed the experiments, analysed the data, and wrote the manuscript. N.G.C. performed and analysed qPCR experiments. R.O. performed tissue culture experiments and quantification. G.Y. helped with statistical analysis and data interpretation. K.N. contributed reagents, and performed data analyses and manuscript revision. F.H.G. contributed to the concept, analysed the data and revised the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.R.M. (muotri@ucsd.edu) or F.H.G. (gage@salk.edu).

# The structural basis for membrane binding and pore formation by lymphocyte perforin

Ruby H. P. Law<sup>1,2\*</sup>, Natalya Lukyanova<sup>3\*</sup>, Ilia Voskoboinik<sup>4,5\*</sup>, Tom T. Caradoc-Davies<sup>6</sup>, Katherine Baran<sup>4</sup>, Michelle A. Dunstone<sup>1,7</sup>, Michael E. D'Angelo<sup>1</sup>, Elena V. Orlova<sup>3</sup>, Fasséli Coulibaly<sup>1</sup>, Sandra Verschoor<sup>4</sup>, Kylie A. Browne<sup>4</sup>, Annette Ciccone<sup>4</sup>, Michael J. Kuiper<sup>8</sup>, Phillip I. Bird<sup>1</sup>, Joseph A. Trapani<sup>4,9\*</sup>, Helen R. Saibil<sup>3\*</sup> & James C. Whisstock<sup>1,2\*</sup>

Natural killer cells and cytotoxic T lymphocytes accomplish the critically important function of killing virus-infected and neoplastic cells. They do this by releasing the pore-forming protein perforin and granzyme proteases from cytoplasmic granules into the cleft formed between the abutting killer and target cell membranes. Perforin, a 67-kilodalton multidomain protein, oligomerizes to form pores that deliver the pro-apoptotic granzymes into the cytosol of the target cell<sup>1–6</sup>. The importance of perforin is highlighted by the fatal consequences of congenital perforin deficiency, with more than 50 different perforin mutations linked to familial haemophagocytic lymphohistiocytosis (type 2 FHL)<sup>7</sup>. Here we elucidate the mechanism of perforin pore formation by determining the X-ray crystal structure of monomeric murine perforin, together with a cryo-electron microscopy reconstruction of the entire perforin pore. Perforin is a thin 'key-shaped' molecule, comprising an amino-terminal membrane attack complex perforin-like (MACPF)/cholesterol dependent cytolyisin (CDC) domain<sup>8,9</sup> followed by an epidermal growth factor (EGF) domain that, together with the extreme carboxy-terminal sequence, forms a central shelf-like structure. A C-terminal C2 domain mediates initial, Ca<sup>2+</sup>-dependent membrane binding. Most unexpectedly, however, electron microscopy reveals that the orientation of the perforin MACPF domain in the pore is inside-out relative to the subunit arrangement in CDCs<sup>10,11</sup>. These data reveal remarkable flexibility in the mechanism of action of the conserved MACPF/CDC fold and provide new insights into how related immune defence molecules such as complement proteins assemble into pores.

The sequence similarity between perforin and complement components C6–C9 of the membrane attack complex strongly suggests that two major branches of the mammalian immune system utilize a pore-forming MACPF fold as the final weapon mediating target cell death<sup>1–6</sup>. Recent structural studies on the non-pore-forming protein Plu-MACPF<sup>8</sup> and the MACPF domain of human complement C8 $\alpha$ <sup>9,12</sup> surprisingly revealed that MACPF proteins are homologous to bacterial CDCs, such as perfringolysin O<sup>8–10,12,13</sup>. In addition to the MACPF domain, perforin contains a Ca<sup>2+</sup>-dependent, membrane-binding C2 domain homologous to the membrane-binding immunoglobulin domain of CDCs<sup>14</sup>. However, without the structures of a complete lytic MACPF protein and a MACPF pore, the mechanisms of perforin function and dysfunction remain unclear. To address these issues, we determined the 2.75-Å-resolution structure of mouse perforin (an oligomerization-impaired variant, R213E)<sup>15</sup> and the cryo-electron microscopy structure of an intact perforin pore.

The perforin monomer structure (Fig. 1a, b, Supplementary Fig. 1 and Supplementary Table 1) roughly resembles the shape of bacterial CDCs, with a long dimension of 125 Å (Supplementary Fig. 2a, b). A

central feature of the perforin MACPF domain is a bent and twisted four-stranded  $\beta$ -sheet flanked by two clusters of  $\alpha$ -helices, termed CH1 and CH2 (Supplementary Fig. 2a–c). In CDCs, the regions equivalent to CH1 and CH2 unwind upon pore formation to insert into membranes as amphipathic  $\beta$ -strands<sup>16,17</sup> (Supplementary Fig. 2d, e). In the perforin monomer, CH1 is loosely held between the central sheet, the C-terminal  $\alpha$ -helix of the MACPF domain and the disulphide constrained EGF-like fold that follows the MACPF domain (Fig. 1a–c). At the end of the EGF domain, a conserved disulphide bond (C407–C241) is formed with the first helix of CH2 (Fig. 1c). The EGF domain is intimately associated with the extreme C-terminal sequence (residues 524–551). Together these structures form a continuous shelf on which the MACPF sits and beneath which hangs a type II (rather than the predicted type I<sup>18</sup>) C2 domain (Fig. 1a, b). Several FHL-associated mutations map to this region (Fig. 1c). The close proximity of the N and C termini of the C2 domain and structural continuity of the shelf region suggest that the C2 domain may have been inserted into an ancestral MACPF protein that contained a C-terminal array of small disulphide constrained structures (Fig. 1c, Supplementary Fig. 3).

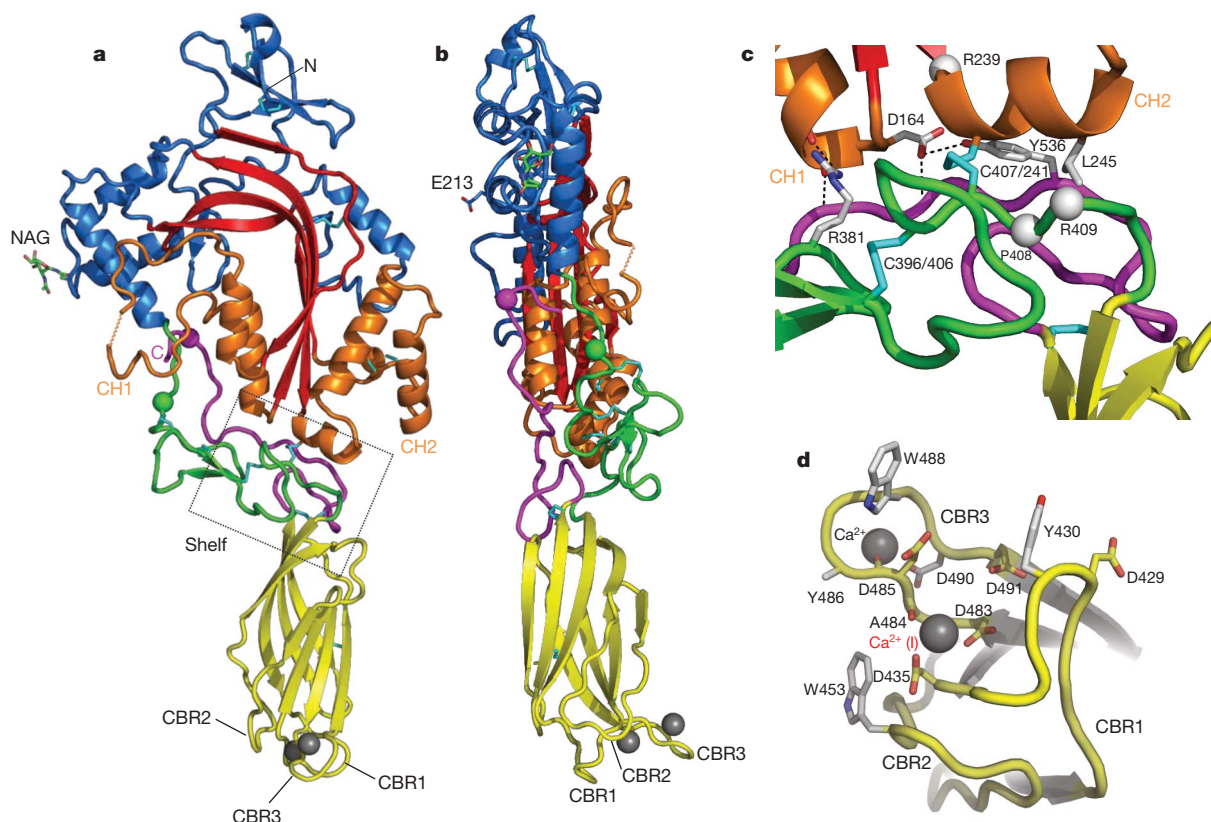
The C2 domain of perforin is important for regulation of its activity; low concentrations of Ca<sup>2+</sup> and acidic pH in the granule prevent premature activation of perforin. On granule exocytosis, higher extracellular Ca<sup>2+</sup> and neutral pH promote membrane binding<sup>1–8,18–21</sup>. The C2 fold can coordinate up to four Ca<sup>2+</sup> atoms (at sites I–IV); these can promote conformational change within the Ca<sup>2+</sup>-binding loops and/or the metal ions themselves may interact with lipid head groups<sup>14,22</sup>. We observed one Ca<sup>2+</sup> atom canonically coordinated in the site I position between the three calcium-binding regions (CBR1–3) of the C2 domain (Fig. 1d). A second Ca<sup>2+</sup> atom is coordinated outside of CBR3 by D490 (Fig. 1d). This site is not conserved in other C2 domains, and D490 is not essential for perforin function<sup>21</sup>. A comparison between the perforin C2 domain with the structures of the apo- and Ca<sup>2+</sup>-bound Munc13-C2B domains<sup>22</sup> reveals that the perforin Ca<sup>2+</sup>-binding site II is unoccupied (Supplementary Fig. 4) and that the functionally important residue D429 (ref. 21) is positioned ~8 Å away from the Ca<sup>2+</sup>-binding sites (Fig. 1d and Supplementary Fig. 4). In apo-Munc13-C2B, D705 (the equivalent residue to D429; Supplementary Fig. 4) is also positioned away from the Ca<sup>2+</sup>-binding sites and shifts on Ca<sup>2+</sup> binding to coordinate the site I and II Ca<sup>2+</sup> atoms<sup>22</sup>. Once both site I and II Ca<sup>2+</sup> atoms are bound, however, the perforin C2 domain will presumably be capable of interacting strongly with membranes, as observed for other C2 family members<sup>22</sup>; indeed several aromatic residues at the C2 base could interact with lipid acyl groups (Fig. 1d).

Following Ca<sup>2+</sup>-mediated membrane interaction, perforin monomers assemble into a pore. To address the mechanism of perforin pore

<sup>1</sup>Department of Biochemistry and Molecular Biology, Monash University, Clayton, Melbourne, Victoria 3800, Australia. <sup>2</sup>The ARC Centre of Excellence in Structural and Functional Microbial Genomics, Monash University, Melbourne, Victoria 3800, Australia. <sup>3</sup>Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, Malet Street, London WC1E 7HX, UK. <sup>4</sup>Cancer Immunology Program, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria 3002, Australia. <sup>5</sup>Department of Genetics, The University of Melbourne, Parkville, Victoria 3010, Australia. <sup>6</sup>Australian Synchrotron, 800 Blackburn Road, Clayton, Melbourne, Victoria 3168, Australia. <sup>7</sup>Department of Microbiology, Monash University, Melbourne, Victoria 3800, Australia. <sup>8</sup>Life Sciences Department, Victorian Partnership of Advanced Computing, Carlton South, Victoria 3053, Australia. <sup>9</sup>Department of Microbiology and Immunology, The University of Melbourne, Parkville, Victoria 3010, Australia.

\*These authors contributed equally to this work.



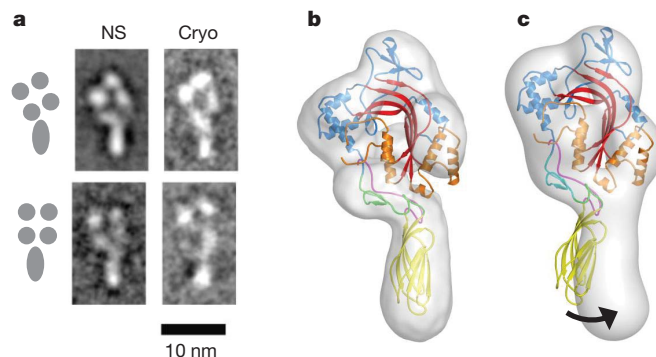


**Figure 1 | Structure of perforin monomers.** **a, b**, Entire monomer, rotated 90°. Red, central sheet of the MACPF domain; orange, CH1 and CH2; blue, the remainder of this domain. Dashed orange line, connectivity around the missing residue P136 in CH1. The perforin MACPF domain superposes with the C8 $\alpha$  MACPF with an r.m.s.d. of 4.46 Å (230 C $\alpha$  atoms; Supplementary Fig. 2c). Green, EGF domain; yellow, C2 domain; magenta, C-terminal region. Shelf region is boxed. Cyan sticks, disulphide bonds (Supplementary Fig. 1); grey spheres, two Ca<sup>2+</sup> atoms. The MACPF domain contains three N-linked oligosaccharides, one of which (attached to N204; in green stick) is visible in electron density (NAG). Green and purple spheres, positions of the two other oligosaccharides (attached to N375 and N548, respectively). In **b**, the position of the R213E mutation is shown in stick. **c**, Cartoon of the shelf region boxed in **a**, with unchanged colour coding, illustrating major interactions formed with the

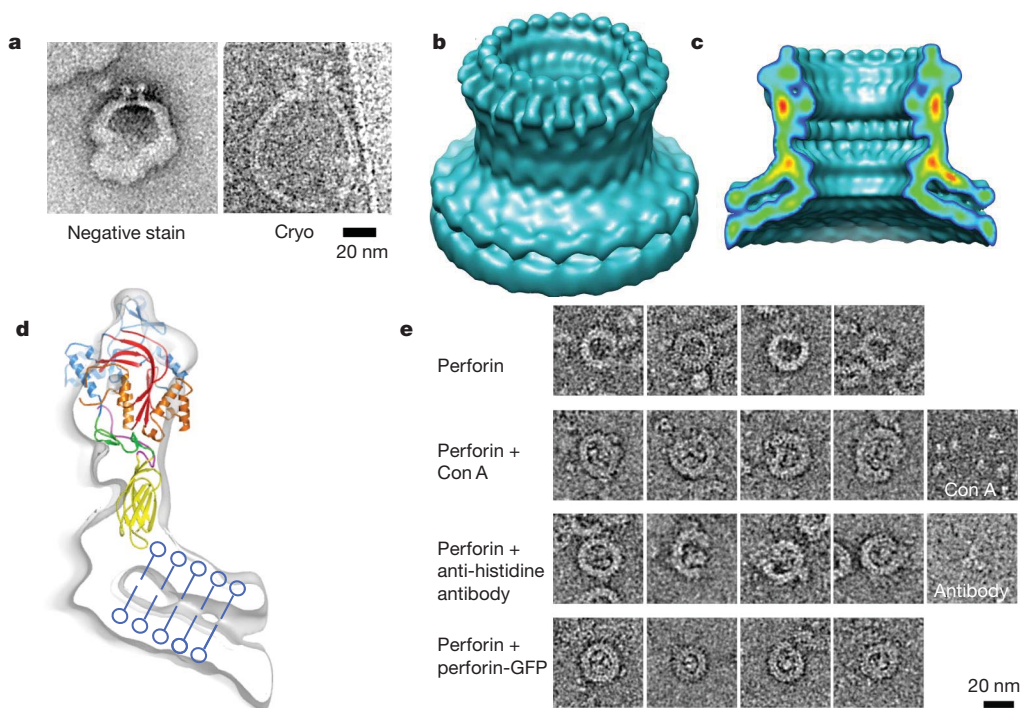
MACPF domain. The sidechains of two residues in the shelf (R381 and Y536) form polar interactions with CH1 and CH2. L245, which is highly conserved, is located at the end of the first helix of CH2 and packs into the shelf. The disulphide bond between the shelf and CH2 (C241/C407) is shown in stick; grey spheres, positions of three residues mutated in human perforin (square bracket) in FHL (P408[P409], R409[R410] and R239[R240]). **d**, Base of the perforin C2 domain. The three CBR loops are in yellow. Ca<sup>2+</sup> atom coordinating residues that are functionally important (as determined by mutagenesis studies<sup>21</sup>) are shown in yellow/red stick. The site I Ca<sup>2+</sup> atom is coordinated by residues D435 and D483 as well as the carbonyl oxygen of A484. A second Ca<sup>2+</sup> atom is located on the other side of CBR3. D429 is located ~8 Å away from the Ca<sup>2+</sup>-binding site, but its position may be influenced by crystal contacts. Four aromatic residues (W453, W488, Y430 and Y486) are shown in grey stick form.

formation, we examined perforin monomers and pores by electron microscopy. Single-particle maps of wild-type mouse perforin monomers obtained with this technique are in good agreement with the perforin crystal structure (Fig. 2a–c). In addition, the images reveal variable angles between the C2 and MACPF domains, suggesting that the shelf region contains a hinge point. In support of this, *B*-factor analysis suggests that the EGF domain (as well as parts of CH1 and CH2) is extremely flexible (Supplementary Fig. 5).

To determine the conformation of perforin in membrane-inserted pores, we recorded electron microscopy images of liposomes containing histidine-tagged, wild-type human perforin pores (Fig. 3a, Supplementary Fig. 6). We used single particle analysis of extracted image regions containing the pores with small surrounding areas of membrane to determine the three-dimensional (3D) structure of the pore (Fig. 3b, c). In marked contrast to the CDCs, in which the monomer undergoes a major collapse and rearrangement in the pore form<sup>10,23</sup>, the perforin monomer was broadly compatible in overall shape and height with the pore profile (Fig. 3d). Most unexpectedly, docking of the perforin crystal structure (minus the CH regions) into the map revealed that the MACPF domain fits significantly better (cross-correlation 0.57) in the orientation opposite to that found in CDC pores (cross-correlation 0.51; Supplementary Fig. 6e, f). However, the



**Figure 2 | Electron microscopy of perforin monomers.** **a**, Averaged images of perforin monomers obtained by classification of different conformations. Schematic views (left), negative stain (NS; middle) and cryo-electron microscopy (Cryo; right) of two conformations. **b, c**, Single-particle negative stain reconstructions of perforin monomer (grey surface), with the crystal structure docked in, showing rotation (arrow) of the C2 domain relative to the 'head' domain.



**Figure 3 | Perforin pore structure.** **a**, Negative stain and cryo-electron microscopy images of liposomes with perforin pores. **b**, **c**, Surface (**b**) and cut-away (**c**) views of a cryo-electron microscopy reconstruction of a perforin pore with 20-fold symmetry. The cut surface is rainbow coloured by density, with red representing the high density regions. The map resolution is 28.5 Å. **d**, Section of the pore map with the perforin crystal structure superposed. Although shown in the monomer conformation, the CH domains would be refolded in

the pore conformation. The C2 domain interacts with the upper leaflet of the membrane bilayer. The lipid bilayer is shown schematically. **e**, Negative stain electron microscopy of perforin pores inserted into lipid monolayers. Perforin is shown alone and with bound Con A, anti-histidine antibody and mixed with a perforin/C-terminal GFP fusion construct. Images of Con A and antibody alone are shown at the right. The labelling results show that both the C terminus and the glycosylation sites face into the pore lumen.

map resolution is limited by a combination of size heterogeneity (Supplementary Fig. 7) and aggregation propensity of perforin pores.

Because of the uncertainties of fitting into a low-resolution map, we performed labelling experiments on mouse perforin pores formed on lipid monolayers<sup>11</sup>. We noted that the perforin N-linked oligosaccharides as well as the C terminus (which includes the histidine tag) all map to the same side of the molecule (Fig. 1a). Accordingly, we used the oligosaccharide-binding lectin concanavalin A (Con A) and a monoclonal antibody to the histidine tag for labelling. In addition, we examined pores formed with a perforin variant C-terminally tagged with green fluorescent protein (GFP). All the results are consistent with the 'inside-out' orientation, in that each probe associated with the interior of the pores (Fig. 3e, Supplementary Fig. 8). Moreover, if the perforin C2 and perfringolysin O immunoglobulin domain structures are superposed (these folds are distantly homologous<sup>24</sup>), their MACPF/CDC domains face in opposite directions (Supplementary Fig. 2). Fitting subunits into the pore density gives a model that is consistent with the flat faces of perforin, which contain complementary charged residues (including R213 on one face and E343 on the other), interacting in the pore form (Fig. 4a–f)<sup>15</sup>.

Despite the homology between CDCs and MACPF proteins and the shared (in the case of perforin) immunoglobulin/C2 membrane-binding domain, the conformation of the monomer in the pore is remarkably different. Sequence alignments reveal amphipathic regions in perforin CH1 and CH2 (Fig. 4e, f and Supplementary Fig. 1), consistent with the hypothesis that, like CDCs, MACPF proteins span membranes via amphipathic  $\beta$ -hairpins<sup>8,9,25</sup> (Supplementary Figs 1, 2). However, in contrast to CDCs, which must buckle to bring CH1 and CH2 close to the membrane surface<sup>10,16,23</sup>, perforin is approximately the same height in the monomer and in the pore structure, suggesting the molecule does not collapse during pore formation (Fig. 3). Accordingly, we note that

the perforin CH1 and CH2 sequences are twice as long as those of their CDC counterparts. The unfurled loops are thus long enough to permit the amphipathic sequences to reach and insert into the membrane (Fig. 4b). Interestingly, because perforin does not open up like CDCs (Fig. 4g–j), the CH1 and CH2 loop must pass over the shelf region (Fig. 4a–d).

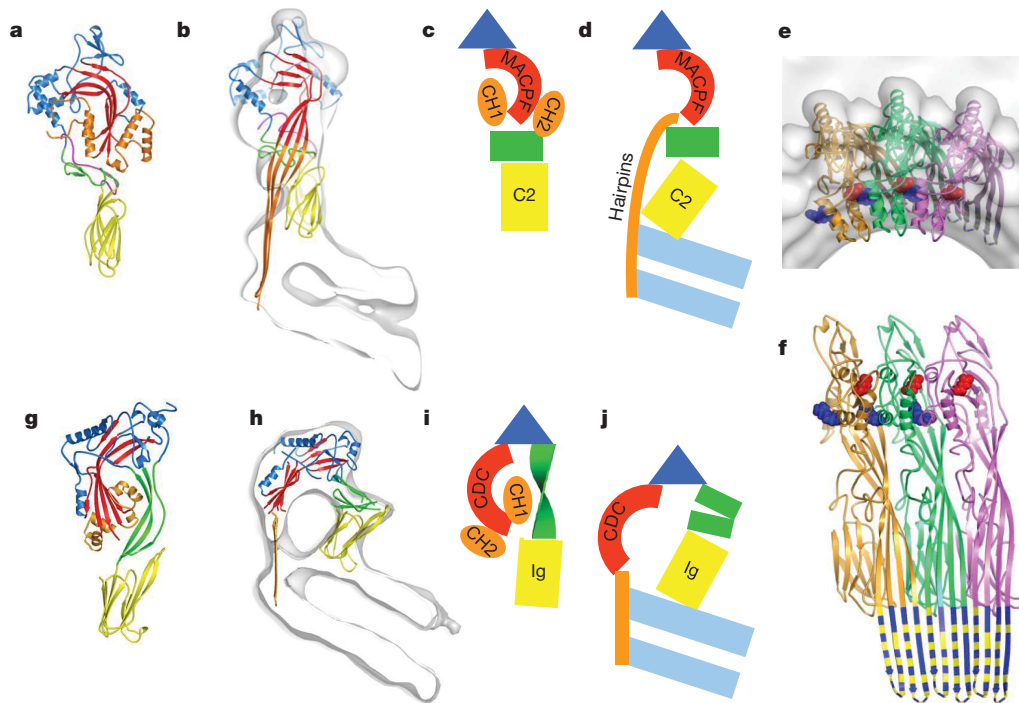
Our electron microscopy analysis revealed a distribution of pore lumen diameters spanning the range 50–300 Å, with the majority formed of 19–24 subunits, corresponding to a lumen of 130–200 Å. Very similar pore features were observed in the membranes of nucleated and non-nucleated cells following attack by intact, minimally stimulated human natural killer cells<sup>26</sup>.

Pore sizes in the range we observe would permit a typical granzyme monomer (50 Å × 50 Å × 45 Å) or indeed a granzyme A dimer (90 Å × 50 Å × 45 Å) to pass readily through the lumen. The pores observed *in vitro* are compatible with their action either in the plasma membrane of the target cell or, as alternatively proposed, in an endosomal membrane after osmotic-stress-induced endocytosis of perforin and granzymes<sup>27</sup>.

Our observation that the perforin pore is lined by oligosaccharides raised the question as to whether these modifications facilitate the delivery of granzymes, some of which bind glycosaminoglycans<sup>28</sup>. We tested this hypothesis, but found that deglycosylated perforin efficiently delivers granzyme B (Supplementary Figs 9 and 10). Furthermore, we note that the glycosylation sites are not conserved in all perforin species (Supplementary Fig. 5), suggesting that this feature is not essential for perforin function.

Finally, an important question is whether the reversed orientation occurs in other members of the MACPF superfamily. Structural studies on the complement C8 $\alpha$  MACPF domain in complex with the lipocalin C8 $\gamma$  reveal that the latter subunit is positioned on the CH2 (convex)





**Figure 4 | Schematic comparison of pore formation in perforin and CDCs.** **a–f**, Perforin. **a**, Monomer (the MACPF domain is in blue, red (central  $\beta$ -sheet) and orange (CH1 and CH2), the EGF domain in green and the C2 domain in yellow). **b**, Model of the pore conformation with the CH regions extended into  $\beta$ -hairpins. After extrusion of the CH regions, some rearrangement is likely around the helical bundle region in the N terminus (blue, upper left), which does not fit inside the density. **c**, **d**, Diagrams of perforin sub-domain organization with the different regions of the molecule represented by coloured shapes corresponding to the colours on the cartoons in **a**, **b**. The lipid bilayer is shown as light blue bars. **e**, **f**, Model of perforin subunit packing in the pore seen

side of the curved sheet<sup>12</sup> (Supplementary Fig. 11). In accordance with previous biochemical and photolabelling studies<sup>29</sup>, a reverse orientation would place C8 $\gamma$  outside the pore to dock the membrane attack complex (MAC) on the membrane. The reverse orientation would also place the CH2 sequence of C8 $\alpha$  and C9 in an appropriate position for interaction with membrane anchored MAC inhibitor CD59 (refs 9, 12). In addition, like perforin, C8 $\alpha$  and C9 both have substantially longer CH1 and CH2 sequences than a typical CDC, consistent with a requirement to span a greater distance to reach the membrane surface. Thus our data suggest that despite their common ancestry<sup>8</sup>, MACPF immune proteins and the bacterial CDCs have undergone an extraordinary structural adaptation to function in opposite orientations.

## METHODS SUMMARY

**Crystallography.** Baculovirus-expressed murine perforin R213E was produced as previously described<sup>15</sup>. Recombinant material was concentrated to 3 mg ml<sup>-1</sup> and crystals obtained in 0.5 M Na acetate, 0.1 M imidazole, pH 6.5. Data from a native compound (Native1) and three heavy atom derivatives (ethylmercury phosphate, ammonium hexachloroiridate(III) and iodine) were collected, and experimental phases (Supplementary Table 1) were obtained by multiple isomorphous replacement with anomalous scattering (MIRAS). Model building was performed using COOT.

**Electron microscopy.** Perforin monomers imaged by negative stain and cryo-electron microscopy were sorted by multivariate statistical analysis and multi-reference alignment into two conformations with different inter-domain angles. Three-dimensional reconstructions were obtained from the negative stain images by a combination of angular reconstitution and projection matching. Negative stain images of pores formed in lipid monolayers<sup>11</sup> were used to determine the symmetry and also for labelling experiments to determine subunit orientation. Three dimensional reconstructions were obtained by angular reconstitution and projection matching from cryo-electron microscopy images of pores formed in liposomes after sorting into different symmetry classes<sup>10</sup>. The perforin crystal

from above (**e**) and from inside (**f**) the pore, showing the interaction of R213 (blue) and E343 (red) in neighbouring molecules<sup>15</sup>. The amphipathic region is coloured by non-hydrophobic (blue) and hydrophobic (yellow) residues. **g–j**, Equivalent views of a CDC. (Panels **g** and **h** are modified from ref. 10.) At the point of protein insertion (**b** and **f**), the membrane is significantly bent. **i**, A twist in the CDC connecting domain (green) reverses the orientation of the pore forming (MACPF/CDC) domain (red) relative to the membrane-surface-binding C2/immunoglobulin (Ig) domain (yellow). The same colour coding is used throughout **a–d** and **g–j**.

structure was manually docked into the electron microscopy maps, and a model of the pore formation was constructed using interactive molecular dynamics to extend CH1 and CH2 to form  $\beta$ -hairpins.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 May; accepted 20 September 2010.

Published online 31 October 2010.

1. Tschopp, J., Masson, D. & Stanley, K. K. Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytotoxicity. *Nature* **322**, 831–834 (1986).
2. Shinkai, Y., Takio, K. & Okumura, K. Homology of perforin to the ninth component of complement (C9). *Nature* **334**, 525–527 (1988).
3. Lichtenheld, M. G. *et al.* Structure and function of human perforin. *Nature* **335**, 448–451 (1988).
4. Lowin, B., Hahne, M., Mattmann, C. & Tschopp, J. Cytolytic T-cell cytotoxicity is mediated through perforin and Fas lytic pathways. *Nature* **370**, 650–652 (1994).
5. Kägi, D. *et al.* Cytotoxicity mediated by T cells and natural killer cells is greatly impaired in perforin-deficient mice. *Nature* **369**, 31–37 (1994).
6. Young, J. D., Cohn, Z. A. & Podack, E. R. The ninth component of complement and the pore-forming protein (perforin 1) from cytotoxic T cells: structural, immunological, and functional similarities. *Science* **233**, 184–190 (1986).
7. Voskoboinik, I., Smyth, M. J. & Trapani, J. A. Perforin-mediated target-cell death and immune homeostasis. *Nature Rev. Immunol.* **6**, 940–952 (2006).
8. Rosado, C. J. *et al.* A common fold mediates vertebrate defense and bacterial attack. *Science* **317**, 1548–1551 (2007).
9. Hadders, M. A., Beringer, D. X. & Gros, P. Structure of C8 $\alpha$ -MACPF reveals mechanism of membrane attack in complement immune defense. *Science* **317**, 1552–1554 (2007).
10. Tilley, S. J., Orlova, E. V., Gilbert, R. J., Andrew, P. W. & Saibil, H. R. Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* **121**, 247–256 (2005).
11. Dang, T. X., Hotze, E. M., Rouiller, I., Tweten, R. K. & Wilson-Kubalek, E. M. Prepore to pore transition of a cholesterol-dependent cytolysin visualized by electron microscopy. *J. Struct. Biol.* **150**, 100–108 (2005).
12. Slade, D. J. *et al.* Crystal structure of the MACPF domain of human complement protein C8 $\alpha$  in complex with the C8 $\gamma$  subunit. *J. Mol. Biol.* **379**, 331–342 (2008).



13. Rossjohn, J., Feil, S. C., McKinstry, W. J., Tweten, R. K. & Parker, M. W. Structure of a cholesterol-binding, thiol-activated cytolyisin and a model of its membrane form. *Cell* **89**, 685–692 (1997).
14. Hurley, J. H. & Misra, S. Signaling and subcellular targeting by membrane-binding domains. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 49–79 (2000).
15. Baran, K. *et al.* The molecular basis for perforin oligomerization and transmembrane pore assembly. *Immunity* **30**, 684–695 (2009).
16. Shepard, L. A. *et al.* Identification of a membrane-spanning domain of the thiol-activated pore-forming toxin *Clostridium perfringens* perfringolysin O: an  $\alpha$ -helical to  $\beta$ -sheet transition identified by fluorescence spectroscopy. *Biochemistry* **37**, 14563–14574 (1998).
17. Shatursky, O. *et al.* The mechanism of membrane insertion for a cholesterol-dependent cytolyisin: a novel paradigm for pore-forming toxins. *Cell* **99**, 293–299 (1999).
18. Urrea Moreno, R. *et al.* Functional assessment of perforin C2 domain mutations illustrates the critical role for calcium-dependent lipid binding in perforin cytotoxic function. *Blood* **113**, 338–346 (2009).
19. Podack, E. R., Young, J. D. & Cohn, Z. A. Isolation and biochemical and functional characterization of perforin 1 from cytolytic T-cell granules. *Proc. Natl Acad. Sci. USA* **82**, 8629–8633 (1985).
20. Young, J. D., Nathan, C. F., Podack, E. R., Palladino, M. A. & Cohn, Z. A. Functional channel formation associated with cytotoxic T-cell granules. *Proc. Natl Acad. Sci. USA* **83**, 150–154 (1986).
21. Voskoboinik, I. *et al.* Calcium-dependent plasma membrane binding and cell lysis by perforin are mediated through its C2 domain: a critical role for aspartate residues 429, 435, 483, and 485 but not 491. *J. Biol. Chem.* **280**, 8426–8434 (2005).
22. Shin, O. H. *et al.* Munc13 C<sub>2</sub>B domain is an activity-dependent Ca<sup>2+</sup> regulator of synaptic exocytosis. *Nature Struct. Mol. Biol.* **17**, 280–288 (2010).
23. Czajkowsky, D. M., Hotze, E. M., Shao, Z. & Tweten, R. K. Vertical collapse of a cytolyisin prepore moves its transmembrane  $\beta$ -hairpins to the membrane. *EMBO J.* **23**, 3206–3215 (2004).
24. Grobler, J. A. & Hurley, J. H. Similarity between C2 domain jaws and immunoglobulin CDRs. *Nature Struct. Biol.* **4**, 261–262 (1997).
25. Ramachandran, R., Tweten, R. K. & Johnson, A. E. The domains of a cholesterol-dependent cytolyisin undergo a major FRET-detected rearrangement during pore formation. *Proc. Natl Acad. Sci. USA* **102**, 7139–7144 (2005).
26. Dourmashkin, R. R., Deteix, P., Simone, C. B. & Henkart, P. Electron microscopic demonstration of lesions in target cell membranes associated with antibody-dependent cellular cytotoxicity. *Clin. Exp. Immunol.* **42**, 554–560 (1980).
27. Thiery, J. *et al.* Perforin activates clathrin- and dynamin-dependent endocytosis, which is required for plasma membrane repair and delivery of granzyme B for granzyme-mediated apoptosis. *Blood* **115**, 1582–1593 (2010).
28. Bird, C. H. *et al.* Cationic sites on granzyme B contribute to cytotoxicity by promoting its uptake into target cells. *Mol. Cell. Biol.* **25**, 7854–7867 (2005).
29. Brickner, A. & Sodetz, J. M. Functional domains of the  $\alpha$  subunit of the eighth component of human complement: identification and characterization of a distinct binding site for the  $\gamma$  chain. *Biochemistry* **24**, 4603–4607 (1985).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** J.C.W. is an Australian Research Council Federation Fellow and Honorary National Health and Medical Research Council of Australia Principal Research Fellow. I.V., F.C. and M.A.D. are NHMRC Career Development Fellows. K.B. is an NHMRC C.J. Martin overseas training fellow. J.A.T. acknowledges the support of an NHMRC Senior Principal Research Fellowship during the course of the work. The authors thank the NHMRC, the ARC, the UK BBSRC and the Wellcome Trust for grant support. We thank the Australian synchrotron beamline scientists for technical support and access to the MX-2 Microfocus Beamline; we thank D. Clare and L. Wang for electron microscopy support, and D. Houldershaw, R. Westlake and K. Mahmood for computing support. We thank D. Steer and the Monash University Proteomics Unit for technical support.

**Author Contributions** R.H.P.L., N.L. and I.V. are joint first authors; J.A.T., H.R.S. and J.C.W. contributed equally to this work. R.H.P.L. crystallized perforin, performed the soaks, collected diffraction data, determined the structure and co-wrote the paper. N.L. performed electron microscopy structural analysis, and co-wrote the paper. I.V. developed the perforin expression system, designed and developed the oligomerization defective variants, produced the perforin variant, co-led the research and co-wrote the paper. T.T.C. collected data and determined the structure, and co-wrote the paper. K.B. developed perforin variants with defective oligomerization. M.A.D. analysed the structure, and co-wrote the paper. M.E.D. performed the bioinformatic research. E.V.O. developed procedures for image processing and analysis. F.C. assisted with determining the structure. S.V., K.A.B. and A.C. produced perforin. M.J.K. performed the modelling experiments. P.I.B. performed bioinformatic experiments, interpreted the data and co-wrote the paper. J.A.T., H.R.S. and J.C.W. analysed the data, led the research and co-wrote the paper.

**Author Information** Structure factors and coordinates are deposited in the Protein Data Bank under accession number 3NSJ. Electron microscopy maps are deposited in the EM Databank (accession numbers EMD-1772 and EMD-1773 for the two conformations of perforin monomer and EMD-1769 for the pore). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.A.T. ([joe.trapani@petermac.org](mailto:joe.trapani@petermac.org)), H.R.S. ([h.saibil@mail.cryst.bbk.ac.uk](mailto:h.saibil@mail.cryst.bbk.ac.uk)) or J.C.W. ([james.whisstock@monash.edu](mailto:james.whisstock@monash.edu)).

## METHODS

**Protein production and crystallography.** Expression and initial purification of recombinant mouse perforin R213E was performed as described<sup>15</sup>, followed by size exclusion chromatography using a HiLoad 16/60 Superdex 200 pg column (GE Healthcare) in a buffer containing 50 mM Tris, 300 mM NaCl, 10% glycerol, 0.05% sodium azide, pH 7.2, plus Complete Protease Inhibitor Cocktail Tablet without EDTA (Roche Applied Science). Purified perforin (3 mg ml<sup>-1</sup>) was crystallized in 0.5 M sodium acetate, 0.1 M imidazole, pH 6.5 at 22 °C. The crystals were flash-cooled in liquid nitrogen using 25% glycerol as the cryoprotectant. All the data sets were collected at the Australian Synchrotron MX2 beamline and were highly anisotropic (as measured by the Diffraction Anisotropy Server; <http://www.doe-mbi.ucla.edu/~sawaya/anisotropy/>)<sup>30</sup>. These data were merged and processed using XDS<sup>31</sup>, POINTLESS and SCALA<sup>32</sup>. Five per cent of the data sets were flagged as a validation set for calculation of the  $R_{\text{free}}$  with neither a  $\sigma$  nor a low-resolution cut-off applied to the data. Experimental phases (Supplementary Table 1) were obtained by the MIRAS method; a native (Native1) data set and three heavy atom derivatives (ethylmercury phosphate, ammonium hexachloroiridate(III) and iodine) were used for phasing. Experimental phasing was carried out using autoSHARP<sup>33</sup>; heavy atom positions were located using SHELXC/SHELXD<sup>34</sup> and refined using SHARP<sup>35</sup> with resulting isomorphous (acentric) and anomalous phasing powers of 0.982 and 0.950, respectively. The initial phases were improved by solvent flipping using SOLOMON<sup>36</sup> and density modification using DM<sup>37</sup>, which dramatically increased the figure of merit (FOM) from 0.34 to 0.86. Such a large increase in FOM is probably due to the very high solvent content of the crystal (70.2%). One molecule was found per asymmetric unit and an initial model was generated using BUCCANEER<sup>38</sup>. Model building was performed using COOT<sup>39</sup> while refinement was performed using PHENIX<sup>40</sup>, REFMAC<sup>41</sup> and autoBUSTER<sup>42</sup>. A higher resolution native data set (Native2, 2.75 Å) was subsequently collected on a crystal soaked in 0.6 M KI and phase extension was carried out to 2.75 Å using DM<sup>37</sup>. Water molecules were added to the model when the  $R_{\text{free}}$  reached 30%.

The recombinant perforin comprises 560 residues; the first 20 amino acids (a signal peptide) were cleaved off during secretion. Residues 21–134 and 136–547 were modelled; P135 in CHI could not be built into density. The model contains two calcium ions, Ca701 and Ca702, that are five and four coordinate, respectively. These were probably scavenged from the environment during expression or purification. Both Ca<sup>2+</sup> ions are in a distorted octahedral geometry. Murine perforin contains three N-linked glycosylation sites, however, density is only observed for the first N-acetylglucosamine attached to N204. The NAG model was made using the PRODRG server (<http://davapc1.bioch.dundee.ac.uk/prodrg/>). The final model also contains three glycerols, two chloride ions and four iodide ions. Crystallographic and structural analysis was performed using CCP4 suite<sup>43</sup>, WHATIF<sup>44</sup> and MUSTANG<sup>45</sup> unless otherwise specified. Figs 1–4 and Supplementary Figs 2, 4, 5, 6 and 11 were generated in part using PYMOL<sup>46</sup>. Structural validation was performed using MolProbity<sup>47</sup>. In the final structure, two residues (L307 and Y486) are in disallowed regions in the Ramachandran plot. The MolProbity score is 1.56, which is in the 100th percentile of structures reported at this resolution. A summary of diffraction and refinement statistics can be found in Supplementary Table 1. The coordinates of perforin, together with the structure factors are deposited in the Protein Data Bank. All diffraction images are deposited in TARDIS (<http://tardis.edu.au/>) and are freely available.

**Electron microscopy sample preparation, data acquisition and preprocessing.** Wild-type mouse and human perforin (which are 68% identical), and mouse perforin C-terminally fused to GFP, were expressed and purified as described<sup>15</sup>.

Mouse perforin monomers were imaged by negative stain (1% uranyl acetate) and cryo-electron microscopy. Human perforin pores were formed on DMPC/cholesterol lipid monolayers as described<sup>11</sup> and imaged by negative staining. Human perforin pores were formed in liposomes as described<sup>10</sup> for pneumolysin at a molar ratio of 1:4,000–1:7,000 protein to lipid in the following buffer: 0.15 M NaCl, 1 mM CaCl<sub>2</sub>, 20 mM HEPES pH 8.0.

Low dose micrographs of negatively stained samples were recorded on Kodak SO163 film using a Tecnai T12 microscope (FEI) at 120 keV and 52,000× magnification. Cryo-electron microscopy images (focal pairs) of perforin monomers were recorded on a Gatan 4k × 4k CCD camera (15 µm per pixel) using a Tecnai Polara microscope (FEI) at 300 keV and 107,000× magnification. Cryo-electron microscopy images of perforin pores in liposomes were collected on a Gatan 4k × 4k CCD camera (15 µm per pixel) on a Tecnai F20 microscope (FEI) at 200 keV and 67,000× magnification.

The defocus and astigmatism of the micrographs were determined with the MRC program CTFIND2<sup>48</sup> and phases were corrected for effects of the contrast transfer function. EMAN/Boxer<sup>49</sup> was used for particle picking.

**Image processing of perforin monomers.** Multivariate statistical analysis (MSA) in Imagic<sup>50</sup> and multi-reference alignment (MRA) using SPIDER<sup>51</sup> were used to

identify and sort two populations of perforin monomers with different angles between the C2 and MACPF domains. A data set of 10,500 negative stain images yielded low-resolution density maps of these two populations by angular reconstitution<sup>50</sup>. Particle orientations were refined in multiple cycles of MRA, MSA and angular reconstitution and the resulting 3D reconstructions were used as initial models for projection matching in SPIDER<sup>51</sup>. The final reconstructions each comprised about 3,400 particles, resulting in structures at 25 and 23 Å resolution estimated by Fourier shell correlation (FSC) with the 0.5 criterion. Comparison with the crystal structure shows that the molecule thickness (25 Å) is broadened in the electron microscopy map by a lack of edge-on views.

**Image processing of perforin pores on lipid monolayers.** Individual images (1,900) of complete oligomeric rings were translationally aligned to their rotationally averaged sum and then classified according to ring diameter using MSA. The class averages were refined by MSA and MRA and their rotational auto-correlations were calculated in Imagic<sup>50</sup> to determine the symmetry. The averaged views of 23- and 28-mer classes presented in Supplementary Fig. 7 each comprised about 120 particles with resolutions of 20.6 and 24.1 Å, respectively, estimated by the 0.5 Fourier ring correlation.

**Image processing of perforin pores in liposomes.** Once perforin pores are formed the liposomes become very unstable and aggregate, making it difficult to collect a large data set. 512 individual images of pore side views were aligned and classified according to their diameters by MSA and MRA as described<sup>52</sup>. After discarding pore views obstructed by contacts with other liposomes and pores, the most populated subset consisted of 94 pore side views with ~16-nm-diameter pores. Class averages with lowest variance and no out of plane tilt along with an average top view of the same diameter were used to obtain low-resolution density maps by angular reconstitution<sup>50</sup> using the range of symmetries C19–C22. After refinement by MRA, MSA and angular reconstitution, a 20-fold 3D reconstruction was chosen as the one with the lowest error. It was used as an initial model for projection matching in SPIDER with up to 20° out of plane tilt. 59 pore views were selected for the final reconstruction, which gave a resolution of 28.5 Å, estimated by 0.5 FSC (Supplementary Fig. 6d).

**Labelling experiments.** Oligomeric rings of mouse perforin were formed on phosphatidylcholine lipid monolayers as described above. After rinsing with buffer, grids were placed for ~5 min on a droplet of either 0.01 mg ml<sup>-1</sup> concanavalin A (Con A, Sigma) or 0.01 mg ml<sup>-1</sup> mouse monoclonal anti-histidine tag antibody (AbD Serotec), rinsed again and stained with 1% uranyl acetate. Inclusion of the lectin-binding reagent  $\alpha$ -methyl D-mannoside (0.1 M) as a control abolished the binding of Con A.

Mouse perforin C-terminally fused to GFP was mixed with wild-type mouse perforin at concentration ratios 1.5:1 and used to form oligomeric rings on monolayers as described above. Micrographs were collected as described above for perforin pores in liposomes. Example fields of the oligomers and ligands are shown in Supplementary Fig. 8.

**Atomic structure fitting.** Manual fitting of atomic coordinates into electron microscopy maps as well as cross-correlation measurements were done using Chimera<sup>53</sup>, which was also used to produce Fig. 3b and c. The handedness shown for the monomer structure was chosen because it provided a slightly better fit. The hand of the low resolution pore map does not affect the modelling of how the flat, key-shaped subunits are packed.

The model of the pore form (Fig. 4b, e and f) was constructed manually using interactive molecular dynamics with NAMD<sup>54</sup> and VMD<sup>54</sup>. The perforin structure was partially constrained to retain major elements of secondary structure. The helical cluster regions were manually unfurled during molecular dynamics simulations to form  $\beta$ -hairpins. The electron microscopy density was used to guide placement of the  $\beta$ -hairpins and the C2 domain.

**Perforin deglycosylation and target cell killing experiments.** Purified wild-type mouse perforin (6–8 µg of 300–350 µg ml<sup>-1</sup> stock) was digested with 1,000 units of Peptide-N-Glycosidase F (PNGaseF; New England Biolabs) under non-denaturing conditions in 45 mM imidazole, 50 mM NaCl and 167 mM Tris-HCl pH 7.4 at 37 °C for 1 h. To verify the efficiency of deglycosylation, we analysed the sample by immunoblotting using untreated and PNGaseF-treated denatured perforin as negative and positive controls, respectively (Supplementary Fig. 9a). Perforin was visualized using rat monoclonal anti-perforin antibody P1-8<sup>55</sup> and secondary polyclonal rabbit anti-rat-horseradish peroxidase antibody. PNGaseF-treated perforin was smaller than untreated controls, suggesting uniform processing of the protein. Mass spectrometry experiments confirmed the size of wild-type untreated perforin (63,548 Da) and perforin deglycosylated under non-denaturing conditions (60,509 Da—the calculated molecular weight of the naked polypeptide is 60,750 Da).

Lytic activity of serially diluted untreated or PNGaseF-treated perforin (under non-denaturing conditions) was tested on Jurkat T-cells using standard <sup>51</sup>Cr release and fluorescence activated cell sorting-based 7-aminoactinomycin D (7-AAD)/

annexin V binding assays (Supplementary Figs 9b and 10, respectively)<sup>56</sup>. PNGaseF treatment had no significant effect on perforin activity compared to the control. From these titrations, we selected perforin concentrations that caused less than 15% Jurkat cell lysis (the amounts varied depending on the type of cytotoxicity assay and the number of cells used), and tested the synergy with increasing concentrations (0.005  $\mu\text{g ml}^{-1}$  to 0.64  $\mu\text{g ml}^{-1}$ ) of purified recombinant granzyme B<sup>57</sup>. The results of <sup>51</sup>Cr release and 7AAD/annexinV assays showed no significant difference between the abilities of control and PNGaseF treated perforin to synergise with granzyme B (Supplementary Figs 9 and 10).

30. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
31. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
32. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
33. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
34. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
35. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
36. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
37. Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72**, 245–270 (1999).
38. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
40. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
41. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
42. Bricogne, G. *et al.* BUSTER, version 2.8.0 (Global Phasing Ltd, Cambridge, UK, 2009).
43. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56 (1990).
45. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. MUSTANG: a multiple structural alignment algorithm. *Proteins* **64**, 559–574 (2006).
46. DeLano, W. L. The PyMOL molecular graphics system. (DeLano Scientific, Palo Alto, 2008); (<http://www.pymol.org>).
47. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–383 (2007).
48. Crowther, R. A., Henderson, R. & Smith, J. M. MRC image processing programs. *J. Struct. Biol.* **116**, 9–16 (1996).
49. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
50. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
51. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
52. White, H. E., Saibil, H. R., Ignatiou, A. & Orlova, E. V. Recognition and separation of single particles with size variation by statistical analysis of their images. *J. Mol. Biol.* **336**, 453–460 (2004).
53. Pettersen, E. F. *et al.* UCSF Chimera — a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
55. Kawasaki, A. *et al.* Perforin, a pore-forming protein detectable by monoclonal antibodies, is a functional marker for killer cells. *Int. Immunol.* **2**, 677–684 (1990).
56. Sutton, V. R. *et al.* Measuring cell death mediated by cytotoxic lymphocytes or their granule effector molecules. *Methods* **44**, 241–249 (2008).
57. Sun, J. *et al.* Expression and purification of recombinant human granzyme B from *Pichia pastoris*. *Biochem. Biophys. Res. Commun.* **261**, 251–255 (1999).



# 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members

Stephane Daffis<sup>1</sup>\*, Kristy J. Szretter<sup>1</sup>\*, Jill Schriewer<sup>5</sup>, Jianqing Li<sup>2</sup>, Soonjeon Youn<sup>1</sup>, John Errett<sup>6</sup>, Tsai-Yu Lin<sup>7</sup>, Stewart Schneller<sup>8</sup>, Roland Zust<sup>9</sup>, Hongping Dong<sup>11</sup>, Volker Thiel<sup>9,10</sup>, Ganes C. Sen<sup>12</sup>, Volker Fensterl<sup>12</sup>, William B. Klimstra<sup>13</sup>, Theodore C. Pierson<sup>7</sup>, R. Mark Buller<sup>4,5</sup>, Michael Gale Jr<sup>4,6</sup>, Pei-Yong Shi<sup>4,11</sup> & Michael S. Diamond<sup>1,2,3,4</sup>

Cellular messenger RNA (mRNA) of higher eukaryotes and many viral RNAs are methylated at the N-7 and 2'-O positions of the 5' guanosine cap by specific nuclear and cytoplasmic methyltransferases (MTases), respectively. Whereas N-7 methylation is essential for RNA translation and stability<sup>1</sup>, the function of 2'-O methylation has remained uncertain since its discovery 35 years ago<sup>2-4</sup>. Here we show that a West Nile virus (WNV) mutant (E218A) that lacks 2'-O MTase activity was attenuated in wild-type primary cells and mice but was pathogenic in the absence of type I interferon (IFN) signalling. 2'-O methylation of viral RNA did not affect IFN induction in WNV-infected fibroblasts but instead modulated the antiviral effects of IFN-induced proteins with tetra-ricopeptide repeats (IFIT), which are interferon-stimulated genes (ISGs) implicated in regulation of protein translation. Poxvirus and coronavirus mutants that lacked 2'-O MTase activity similarly showed enhanced sensitivity to the antiviral actions of IFN and, specifically, IFIT proteins. Our results demonstrate that the 2'-O methylation of the 5' cap of viral RNA functions to subvert innate host antiviral responses through escape of IFIT-mediated suppression, and suggest an evolutionary explanation for 2'-O methylation of cellular mRNA: to distinguish self from non-self RNA. Differential methylation of cytoplasmic RNA probably serves as an example for pattern recognition and restriction of propagation of foreign viral RNA in host cells.

Most eukaryotic mRNA contains a 5' Cap 0 (7mGpppN) structure with a methyl group at the N-7 position. In higher eukaryotes, methylation of cellular mRNA occurs additionally at the 2'-O site of the penultimate (7mGpppNm, Cap 1) and antepenultimate (7mGpppNmNm, Cap 2) 5' nucleotides in the nucleus and cytoplasm, respectively<sup>3,5</sup>. Many viral mRNAs also contain Cap 1 and 2 structures, but cap acquisition occurs distinctly among virus families<sup>2,6</sup>. RNA and DNA viruses that replicate in the cytoplasm cannot use the host nuclear capping machinery, and thus have evolved MTases to facilitate N-7 and 2'-O capping or mechanisms to 'snatch' the cap from host cell mRNA<sup>1</sup>. It remains unclear how 2'-O methylation contributes to viral infection or cellular mRNA homeostasis<sup>2,3</sup>.

*Flavivirus* is a genus of positive-strand RNA viruses with a 5' Cap 1 structure that is generated by an MTase in the NS5 protein<sup>7</sup>. Whereas mutations abrogating the N-7 MTase activity abort WNV infection, an E218A substitution that completely abolished the 2'-O but not N-7 MTase activity (Supplementary Fig. 1) did not affect replication in permissive BHK cells<sup>8</sup>. Although C57BL/6 mice infected subcutaneously with the parental WNV wild-type (WNV-WT) strain had an approximately 40% mortality rate, recipients of WNV-E218A showed

0% mortality, even at high challenge doses (Fig. 1a,  $P < 0.05$ ,  $n = 10$ ) or after direct intracranial infection (Fig. 1c). Levels of WNV-E218A after subcutaneous inoculation were markedly decreased in the spleen, serum or brain compared with infection by WNV-WT (Fig. 1b).

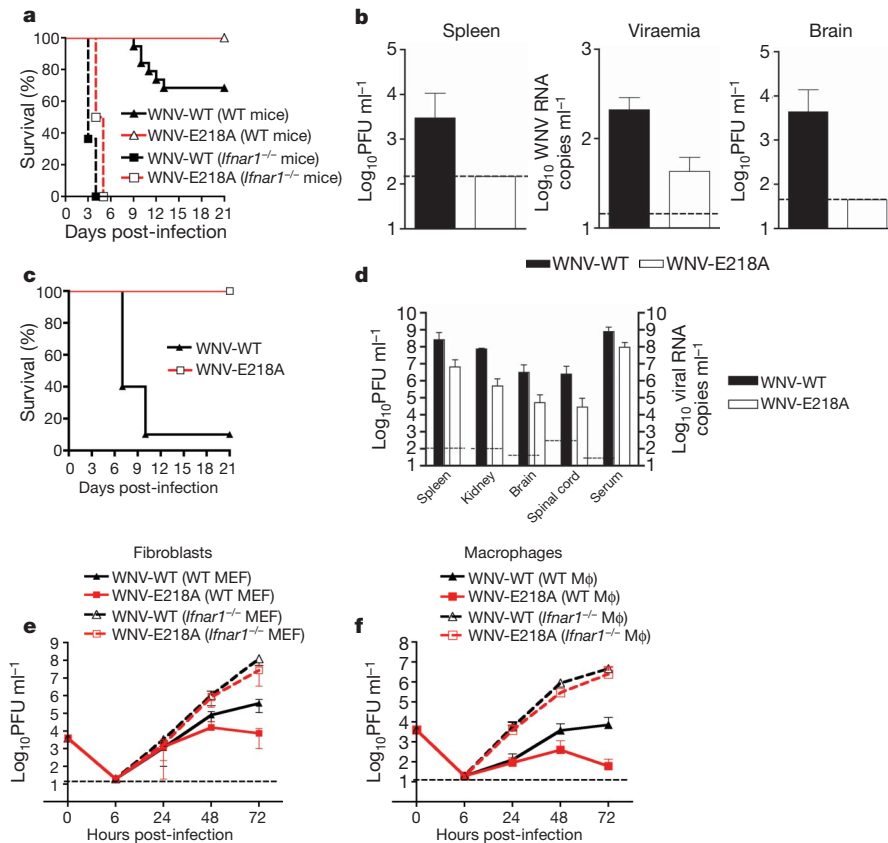
Because dissemination of WNV-E218A was aborted *in vivo*, we assessed whether 2'-O methylation restricted the protective IFN-induced immune response. Mice lacking type I IFN signalling (*Ifnar1*<sup>-/-</sup>) that were infected with WNV-WT showed 100% mortality and a mean time to death of 3.5 days, as seen previously<sup>9</sup> (Fig. 1a). Remarkably, *Ifnar1*<sup>-/-</sup> mice infected with the WNV-E218A exhibited a similar phenotype with only a slightly delayed mean time to death of 4.5 days. *Ifnar1*<sup>-/-</sup> mice infected with WNV-E218A at day 3 sustained tissue titres that approached those of WNV-WT (Fig. 1d). Thus 2'-O methylation of WNV RNA is required for virulence *in vivo*, and its absence renders the virus sensitive to the IFN response.

Analysis of viral growth in primary mouse embryonic fibroblasts (MEFs) and macrophages (Mφ), which both produce and respond to type I IFN after WNV infection<sup>10</sup>, confirmed attenuation of WNV-E218A in wild-type cells (50- and 151-fold lower at 72 h,  $P < 0.05$ ,  $n = 3$  in MEF and Mφ, respectively) and restored growth in *Ifnar1*<sup>-/-</sup> cells (Fig. 1e, f). Replication of WNV-E218A was also rescued in *Irf3*<sup>-/-</sup>, *Irf3*<sup>-/-</sup> × *Irf7*<sup>-/-</sup> or *IPS-1*<sup>-/-</sup> cells that had altered or abolished IFN-α/β responses<sup>11</sup> (Supplementary Figs 2a-c and 3a-d, respectively), but not in *Irf7*<sup>-/-</sup> or *Thr3*<sup>-/-</sup> cells, which have normal IFN-β or IFN-α and -β responses after WNV infection, respectively<sup>10,12</sup> (Supplementary Fig. 2d, e). These experiments confirmed that rescue of WNV-E218A in primary cells requires attenuation of the IFN response.

Because 2'-O methylation rendered WNV-WT less susceptible to the IFN response than WNV-E218A, we hypothesized that it might directly limit IFN induction by affecting the avidity of viral RNA for the host sensor, RIG-I. However, direct binding assays with recombinant RIG-I and 2'-O unmethylated or methylated WNV RNA (5' untranslated region) showed no change in binding (Supplementary Fig. 4). It remained possible that 2'-O methylation of WNV RNA affected other proteins required for transcriptional activation of the IFN-β gene. To evaluate this idea, *Ifnar1*<sup>-/-</sup> MEFs, which produce IFN-β without responding to it, were infected at a high multiplicity of infection (MOI) and IFN-β mRNA was measured. Notably, both WNV-WT and WNV-E218A stimulated IFN-β transcription equivalently after infection (Fig. 2a). Thus a lack of 2'-O methylation does not affect pathogen sensing or IFN induction. To address whether 2'-O methylation of viral RNA serves to antagonize or evade IFN effector functions, *IPS-1*<sup>-/-</sup> MEFs, which do not produce type I IFN after WNV infection but can respond to it<sup>11</sup>, were exposed to IFN-β to

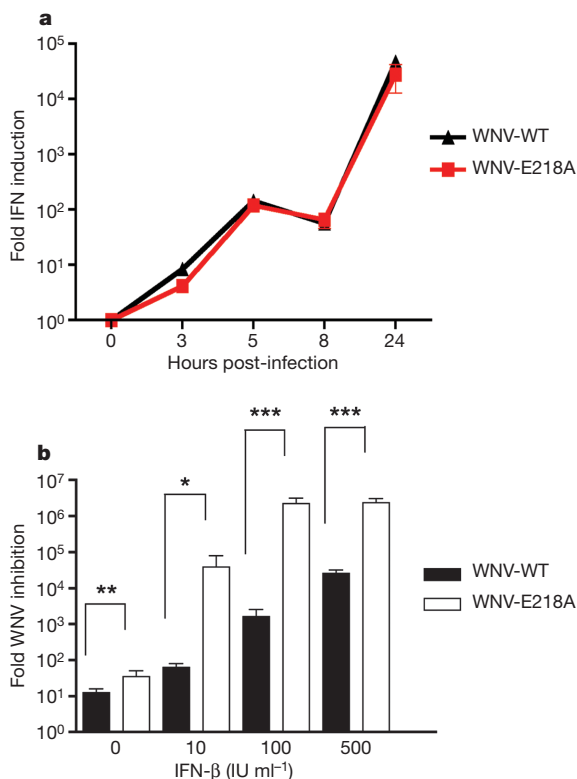
<sup>1</sup>Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>2</sup>Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>3</sup>Department of Pathology & Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>4</sup>Midwest, Pacific Northwest, and Northeast Regional Centers for Biodefense and Emerging Infectious Diseases Research. <sup>5</sup>Department of Molecular Microbiology and Immunology, Saint Louis University School of Medicine, St Louis, Missouri 63104, USA. <sup>6</sup>Department of Immunology, University of Washington School of Medicine, Seattle, Washington 98195-7650, USA. <sup>7</sup>Viral Pathogenesis Section, Laboratory of Viral Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>8</sup>Department of Chemistry and Biochemistry, Auburn University, Auburn, Alabama 36849-5312, USA. <sup>9</sup>Institute of Immunobiology, Kanton Hospital St Gallen, 9007 St Gallen, Switzerland. <sup>10</sup>Vetsuisse Faculty, University of Zürich, 8006 Zürich, Switzerland. <sup>11</sup>Wadsworth Center, New York State Department of Health, Albany, New York 12208, USA. <sup>12</sup>Cleveland Clinic, Lerner Research Institute, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. <sup>13</sup>Department of Microbiology and Molecular Genetics, University of Pittsburgh Medical School, 3501 Fifth Avenue, Pittsburgh, Pennsylvania 15261, USA.

\*These authors contributed equally to this work.



**Figure 1** | WNV-E218A is attenuated in wild-type mice and cells but is virulent in *Ifnar1*<sup>-/-</sup> mice and cells. **a**, Survival curves of wild-type and *Ifnar1*<sup>-/-</sup> C57BL/6 mice after subcutaneous infection with WNV-WT or WNV-E218A. **b**, Virus replication in wild-type mice in blood (day 4), spleen (day 4) or brain (day 8) after subcutaneous infection with WNV-WT or WNV-E218A. **c**, Survival curves of wild-type mice after intracranial infection with

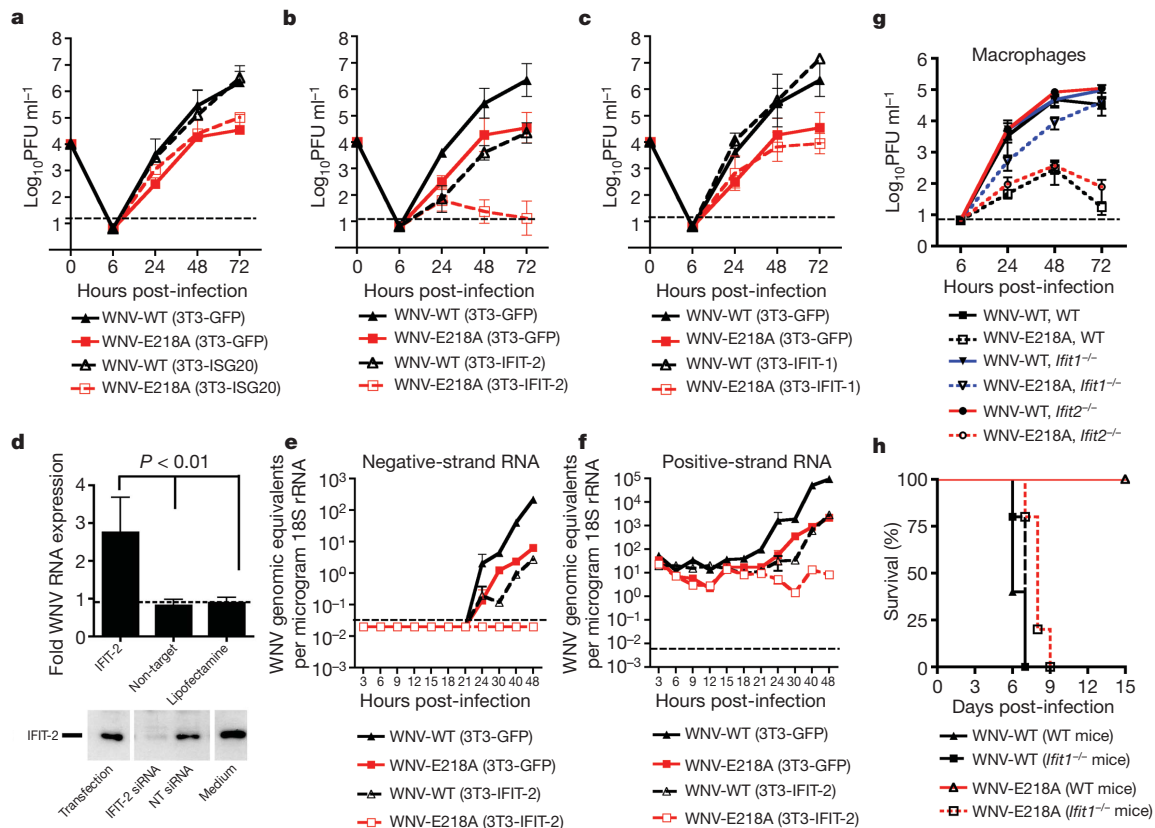
WNV-WT ( $10^1$ ) or WNV-E218A ( $10^5$  plaque-forming units (PFU)). **d**, Viral burden in the serum, spleen, kidney, spinal cord and brain from *Ifnar1*<sup>-/-</sup> mice at day 3 after infection. **e, f**, Replication of WNV-WT and WNV-E218A in wild-type or *Ifnar1*<sup>-/-</sup> MEFs (**e**) or MΦ (**f**). Results are the average of three experiments performed in triplicate. Error bars, s.d.; dashed line, limit of sensitivity of the assay.



induce ISGs, and then infected. WNV-E218A displayed increased sensitivity to IFN- $\beta$  pretreatment compared with WNV-WT (2,400,000- and 20,000-fold inhibition with 500 international units  $\text{ml}^{-1}$  of IFN- $\beta$ , respectively) (Fig. 2b).

IFN induces hundreds of ISGs, some of which may have antiviral effector functions<sup>13</sup>. Among these, *Ifit* family members (for example, *Ifit1* and *Ifit2* (also known as *ISG56* and *ISG54*, respectively)) are induced after WNV infection<sup>14</sup>, reduced in *Irf3*<sup>-/-</sup> and *Ifnar1*<sup>-/-</sup> cells (ref. 15 and Supplementary Fig. 5) and inhibit replication of some viruses<sup>16–18</sup> in part, by interacting with eIF3 and limiting translation of viral mRNA<sup>19,20</sup>. To assess whether differential 2'-O methylation of viral RNA might affect suppression by IFIT-1 and/or IFIT-2, we evaluated infection in 3T3 MEFs expressing a murine *Ifit1* or *Ifit2* transgene. As observed in primary cells, WNV-E218A replication in control 3T3 cells was reduced (~5- to 60-fold decrease at 24–72 h,  $P < 0.05$ ,  $n = 3$ ) compared with WNV-WT, confirming that 2'-O methylation is required for optimal infectivity (Fig. 3a). Transgenic expression of IFIT-2 reduced infection of WNV-WT (~56- to 100-fold decrease at 24–72 h,  $P < 0.0005$ ,  $n = 3$ ) (Fig. 3b) compared with replication in 3T3-green fluorescent protein (GFP) cells. In comparison, expression of IFIT-2

**Figure 2** | 2'-O methylation of viral RNA alters the sensitivity of WNV to the antiviral effects of IFN. **a**, IFN- $\beta$  gene induction in *Ifnar1*<sup>-/-</sup> MEF after WNV-WT or WNV-E218A infection. Results are representative of three independent experiments performed in duplicate. **b**, Viral replication in *IPS-1*<sup>-/-</sup> MEF after IFN- $\beta$  pretreatment. The data are the average of two independent experiments performed in triplicate, and the asterisks indicate differences that are statistically significant (\*\* $P < 0.005$ ; \*\*\* $P < 0.0001$ ; \* $P < 0.05$ ). Error bars, s.d. IU, international units.



**Figure 3 | WNV-E218A is more sensitive to the antiviral actions of *Ifit* genes.** **a–c**, Viral replication of WNV-WT or WNV-E218A in 3T3 MEFs transgenically expressing GFP (**a–c**), ISG20 (**a**), IFIT-2 (**b**) or IFIT-1 (**c**). The data are the average of three experiments performed in duplicate. **d**, siRNA knockdown of IFIT-2 enhances replication of WNV-E218A. 3T3 cells were transfected with a non-target (NT) or IFIT-2 siRNA and then infected with WNV-E218A. One day post-infection cells were collected and (top) viral RNA was assayed by quantitative reverse transcriptase PCR. The data are the average

virtually abolished replication of WNV-E218A (up to 2,700-fold decrease at 72 h,  $P < 0.0005$ ,  $n = 3$ ) (Fig. 3b). Expression of IFIT-1 in 3T3 cells had minimal inhibitory effects on WNV infection (Fig. 3c). To confirm the linkage between IFIT-2 expression and restriction of infection, short interfering RNA (siRNA) knockdown experiments were performed. Transfection of a sequence-specific siRNA that reduced protein expression of IFIT-2 enhanced replication of WNV-E218A ( $P < 0.01$ ,  $n = 3$ ) (Fig. 3d). These experiments demonstrate that mouse IFIT-2 is an antiviral effector of IFN actions, whose inhibitory activity is minimized by 2'-O methylation of viral RNA.

Although IFIT family orthologues exist over a broad evolutionary time-frame<sup>21</sup>, humans have a distinct complement of *Ifit* genes (*Ifit1* (ISG56), *Ifit2* (ISG54), *Ifit3* (ISG60) and *Ifit5* (ISG58)). Transient transgenic expression of human IFIT-5 but not IFIT-1, IFIT-2 or IFIT-3 in human 293T cells inhibited infection of WNV-E218A ( $P = 0.003$ ,  $n = 3$ ) (Supplementary Fig. 6), which suggests a species-specificity of *Ifit* genes in restricting WNV lacking 2'-O methylated RNA.

We assessed the stage of the WNV life cycle that was restricted by mouse IFIT-2. Using strand-specific quantitative reverse transcriptase PCR to quantify genomic (positive strand) and replicative intermediate (negative strand) viral RNA, we found that in control 3T3 cells each increased by 18 h after infection (Fig. 3e, f), whereas the expression of mouse IFIT-2 delayed production of both by approximately 15 h in the context of WNV-WT infection. In comparison, increases in negative and positive strand RNA were abolished in IFIT-2 transgenic cells infected with WNV-E218A. The levels of WNV-E218A positive-strand RNA remained essentially constant over the time course,

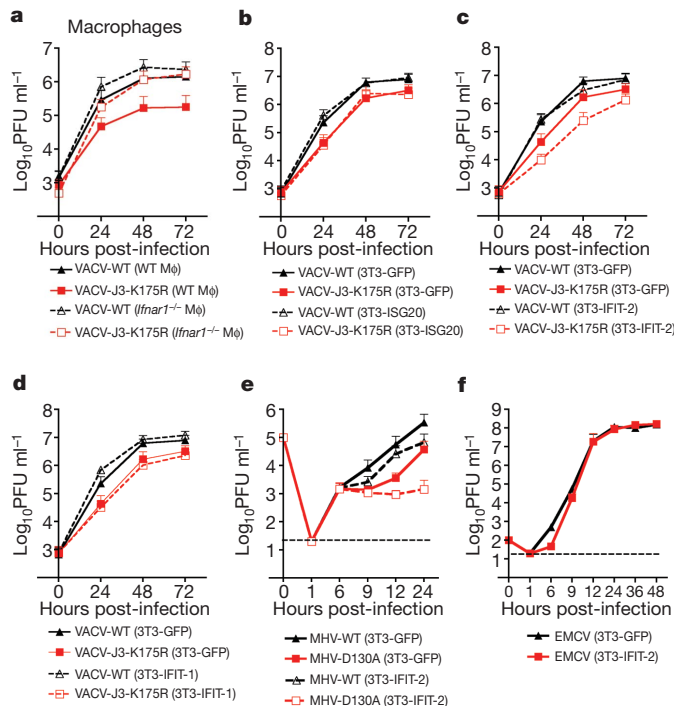
of three experiments performed in duplicate. Bottom, knockdown of IFIT-2 protein was confirmed by western blot. **e, f**, Murine IFIT-2 expression prevents accumulation of negative- and positive-strand viral RNA in WNV-E218A-infected cells. **g**, Replication of WNV-E218A is attenuated in wild-type and *Ifit2*<sup>-/-</sup> Mφ but restored in *Ifit1*<sup>-/-</sup> cells. **h**, Survival curves of wild-type or *Ifit1*<sup>-/-</sup> mice after intracranial challenge with 10<sup>5</sup> plaque-forming units of WNV-WT or WNV-E218A. Error bars, s.d.; dashed line, limit of sensitivity of the assay.

suggesting that the lack of 2'-O methylation did not affect viral RNA stability. Thus mouse IFIT-2 blocks infection of the E218A mutant in fibroblasts at or before negative-strand synthesis.

As other virus families encode 2'-O MTases, we sought to determine if 2'-O-methylation-dependent evasion of IFIT proteins functions as a more general immune escape mechanism. We obtained a vaccinia virus (VACV) mutant (J3-K175R) that lacked 2'-O MTase activity, replicated normally in BSC40 cells<sup>22</sup> but was attenuated in wild-type Mφ (approximately six- to eightfold reduction at 24–72 h) and fully rescued in *Ifnar1*<sup>-/-</sup> Mφ (Fig. 4a). Growth curves with VACV-WT and VACV-J3-K175R in 3T3 cells expressing GFP or ISG20 confirmed an essential role of 2'-O methylation in poxvirus infection (approximately three- to fivefold reduction at 24–72 h,  $P < 0.005$ ,  $n = 3$ ) (Fig. 4b). Transgenic expression of IFIT-2, however, did not affect replication of VACV-WT ( $P > 0.5$ ,  $n = 3$ ), which suggests that IFIT-2 lacks activity against VACV-WT or that the virus efficiently antagonizes its antiviral effect. Expression of mouse IFIT-2 but not IFIT-1 further reduced infection of VACV-J3-K175R (6- to 25-fold decrease,  $P < 0.01$ ,  $n = 3$ ) (Fig. 4c, d). Consistent with these findings, wild-type C57BL/6 mice were resistant to lethal challenge with VACV-J3-K175R (0% lethality,  $n = 6$ ) but sensitive to infection with VACV-WT (100% lethality,  $n = 13$ ). In contrast, in *Ifnar1*<sup>-/-</sup> mice, VACV-J3-K175R was virulent as all animals succumbed to infection with similar kinetics compared with those infected with VACV-WT (Supplementary Fig. 7).

We examined the replication of a wild type and 2'-O MTase mutant (D130A in the nsp16 protein)<sup>23</sup> of mouse hepatitis virus (MHV). MHV-D130A was more sensitive to the effects of IFN-β pretreatment





**Figure 4 | Poxvirus and coronavirus mutants lacking 2'-O methylation are more sensitive to the antiviral effects of murine IFIT-2.** a–d, Studies with VACV. a, Viral replication of VACV-WT or VACV-J3-K175R in wild-type or *Ifnar1*<sup>−/−</sup> Mφ (a) or 3T3 MEF expressing GFP (b–d), ISG20 (b), *Ifit2* (c) or *Ifit1* (d). e, Viral replication of MHV-WT or MHV-D130A in 3T3 cells expressing GFP or IFIT-2. f, Viral replication of EMCV in 3T3 cells expressing GFP or IFIT-2. Error bars, s.d.; dashed line, limit of sensitivity of the assay.

(Supplementary Fig. 8), attenuated in control 3T3 cells (approximately 6- to 15-fold reduction at 9–24 h,  $P < 0.05$ ,  $n = 3$ ) (Fig. 4e), and sensitivity to transgenic expression of mouse IFIT-2 (approximately 8- to 234-fold reduction,  $P < 0.05$ ,  $n = 3$ ) compared with MHV-WT (approximately two- to fivefold decrease at 9–24 h,  $P < 0.05$ ,  $n = 3$ ). Thus, analogous to flaviviruses and poxviruses, the 2'-O methylation of coronavirus RNA supports evasion from the antiviral effects of IFIT-2. In contrast, transgenic expression of IFIT-2 did not affect replication of a picornavirus, which lacks a 5' cap structure (Fig. 4f).

To confirm the role of IFIT proteins in restricting viruses lacking 2'-O methylation, growth curves were performed in wild-type, *Ifit1*<sup>−/−</sup> or *Ifit2*<sup>−/−</sup> Mφ. Surprisingly, the infectivity of WNV-E218A was almost completely rescued in IFIT-1<sup>−/−</sup> Mφ (2,300-fold increase in titre at 72 h,  $P < 0.04$ ) but not in *Ifit2*<sup>−/−</sup> Mφ (Fig. 3g), and the virulence of WNV-E218A was almost entirely restored in *Ifit1*<sup>−/−</sup> mice (Fig. 3h). Thus, in primary Mφ and in mice, IFIT-1 plays a dominant role in restricting infection of WNV lacking 2'-O methylation.

We demonstrate that among unrelated RNA and DNA viruses that replicate in the cytoplasm and contain 5' cap structures, 2'-O methylation of viral RNA enhances virulence through evasion of intrinsic cellular defence mechanisms. 2'-O methylation of cellular RNA may have evolved as a means of distinguishing self from non-self RNA by the host during virus infection. Induction of *Ifit* family genes, several of which attenuate translation<sup>19,20,24</sup>, could preferentially recognize viral mRNA lacking 2'-O methylation and selectively restrict propagation. Plants, which lack an IFN response network or *Ifit* family member orthologues, and their viruses, accordingly lack 2'-O-methylation of mRNA. Given that host 2'-O methylation of cellular mRNA largely occurs in the nucleus, pharmacological strategies that disrupt cytoplasmic 2'-O MTase activity could represent a novel class of therapy against several globally relevant pathogenic viruses that replicate exclusively in the cytoplasm.

## METHODS SUMMARY

**Viruses.** WNV-WT and WNV-E218A were propagated in BHK21 cells as described<sup>8</sup>. VACV-WT and VACV-J3-K175R<sup>22</sup> (a gift from R. Condit) and encephalomyocarditis virus (EMCV) (strain K) were propagated in HeLa and L929 cells, respectively. Generation of MHV-WT (strain A59) and MHV-D130A recombinant coronaviruses has been described<sup>25</sup>.

**Mouse experiments.** C57BL/6 wild-type and immunodeficient (*Ifnar1*<sup>−/−</sup>, *Ifit1*<sup>−/−</sup>, *Ifit2*<sup>−/−</sup>, *Irf3*<sup>−/−</sup>, *Irf7*<sup>−/−</sup>, *Irf3*<sup>−/−</sup> × *Irf7*<sup>−/−</sup> and *IPS-1*<sup>−/−</sup>) mice were bred at Washington University. Infection experiments were performed with approval of the Washington University and St Louis University Animal Studies Committees. Viral titres in blood and organs were quantified as previously described<sup>11</sup>.

**Cell culture and viral infection.** Bone-marrow-derived Mφ and MEF were generated as described<sup>11</sup>. 3T3 fibroblasts expressing GFP or ISG were previously described<sup>18</sup>. Cells were infected with WNV, VACV, MHV or EMCV at MOIs of 0.01, 1, 1 and 0.001, respectively. Lysates or supernatants were titred by plaque assay on BHK21-15 cells for WNV and EMCV, BSC-1 cells for VACV and L929 cells for MHV.

**Quantification of IFN-β mRNA.** *Ifnar1*<sup>−/−</sup> MEFs were infected at an MOI of 10 with WNV-WT or WNV-E218A. Total RNA was isolated, treated with DNase (Qiagen), and IFN-β mRNAs were amplified by quantitative reverse transcriptase PCR as described previously<sup>11</sup>.

**IFN-β pretreatment experiment.** *IPS-1*<sup>−/−</sup> MEFs were pretreated with increasing doses of mouse IFN-β (PBL Laboratories) for 24 h and then infected with WNV or MHV at an MOI of 0.1. Supernatants were collected at 48 or 12 h after infection, respectively, and titred by plaque assay.

**Strand-specific real-time reverse transcriptase PCR.** Quantification of positive- and negative-strand WNV RNA was performed using a T7-tagged primer strategy<sup>9</sup>. Fibroblasts expressing GFP or mouse IFIT-2 were infected with WNV-WT or WNV-E218A at an MOI of 1 and total RNA was collected at indicated time points.

Received 24 February; accepted 13 September 2010.

- Furuichi, Y. & Shatkin, A. J. Viral and cellular mRNA capping: past and prospects. *Adv. Virus Res.* **55**, 135–184 (2000).
- Wei, C. M. & Moss, B. Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc. Natl Acad. Sci. USA* **72**, 318–322 (1975).
- Wei, C. M., Gershowitz, A. & Moss, B. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell* **4**, 379–386 (1975).
- Muthukrishnan, S., Moss, B., Cooper, J. A. & Maxwell, E. S. Influence of 5'-terminal cap structure on the initiation of translation of vaccinia virus mRNA. *J. Biol. Chem.* **253**, 1710–1715 (1978).
- Langberg, S. R. & Moss, B. Post-transcriptional modifications of mRNA. Purification and characterization of cap I and cap II RNA (nucleoside-2'-)-methyltransferases from HeLa cells. *J. Biol. Chem.* **256**, 10054–10060 (1981).
- Fechter, P. & Brownlee, G. G. Recognition of mRNA cap structures by viral and cellular proteins. *J. Gen. Virol.* **86**, 1239–1249 (2005).
- Dong, H., Zhang, B. & Shi, P. Y. Flavivirus methyltransferase: a novel antiviral target. *Antiviral Res.* **80**, 1–10 (2008).
- Zhou, Y. et al. Structure and function of flavivirus NS5 methyltransferase. *J. Virol.* **81**, 3891–3903 (2007).
- Samuel, M. A. & Diamond, M. S. Alpha/beta IFN protects against lethal West Nile virus infection by restricting cellular tropism and enhancing neuronal survival. *J. Virol.* **79**, 13350–13361 (2005).
- Daffis, S. et al. Interferon regulatory factor IRF-7 induces the antiviral alpha interferon response and protects against lethal West Nile virus infection. *J. Virol.* **82**, 8465–8475 (2008).
- Daffis, S., Suthar, M. S., Szretter, K. J., Gale, M. Jr & Diamond, M. S. Induction of IFN-β and the innate antiviral response in myeloid cells occurs through an IPS-1-dependent signal that does not require IRF-3 and IRF-7. *PLoS Pathog.* **5**, e1000607 (2009).
- Daffis, S., Samuel, M. A., Suthar, M. S., Gale, M. Jr & Diamond, M. S. Toll-like receptor 3 has a protective role against West Nile virus infection. *J. Virol.* **82**, 10349–10358 (2008).
- Der, S. D., Zhou, A., Williams, B. R. & Silverman, R. H. Identification of genes differentially regulated by interferon α, β, or γ using oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **95**, 15623–15628 (1998).
- Wacher, C. et al. Coordinated regulation and widespread cellular expression of interferon-stimulated genes (ISG) ISG-49, ISG-54, and ISG-56 in the central nervous system after infection with distinct viruses. *J. Virol.* **81**, 860–871 (2007).
- Daffis, S., Samuel, M. A., Keller, B. C., Gale, M. Jr & Diamond, M. S. Cell-specific IRF-3 responses protect against West Nile virus infection by interferon-dependent and -independent mechanisms. *PLoS Pathog.* **3**, e106 (2007).
- Fensterl, V., White, C. L., Yamashita, M. & Sen, G. C. Novel characteristics of the function and induction of murine p56 family proteins. *J. Virol.* **82**, 11045–11053 (2008).
- Sumpter, R. Jr et al. Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J. Virol.* **79**, 2689–2699 (2005).
- Zhang, Y., Burke, C. W., Ryman, K. D. & Klimstra, W. B. Identification and characterization of interferon-induced proteins that inhibit alphavirus replication. *J. Virol.* **81**, 11246–11255 (2007).

19. Terenzi, F., Hui, D. J., Merrick, W. C. & Sen, G. C. Distinct induction patterns and functions of two closely related interferon-inducible human genes, ISG54 and ISG56. *J. Biol. Chem.* **281**, 34064–34071 (2006).
20. Hui, D. J., Terenzi, F., Merrick, W. C. & Sen, G. C. Mouse p56 blocks a distinct function of eukaryotic initiation factor 3 in translation initiation. *J. Biol. Chem.* **280**, 3433–3440 (2005).
21. Sarkar, S. N. & Sen, G. C. Novel functions of proteins encoded by viral stress-inducible genes. *Pharmacol. Ther.* **103**, 245–259 (2004).
22. Latner, D. R., Thompson, J. M., Gershon, P. D., Storrs, C. & Condit, R. C. The positive transcription elongation factor activity of the vaccinia virus J3 protein is independent from its (nucleoside-2'-O-) methyltransferase and poly(A) polymerase stimulatory functions. *Virology* **301**, 64–80 (2002).
23. Decroly, E. *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'-O-) methyltransferase activity. *J. Virol.* **82**, 8071–8084 (2008).
24. Hui, D. J., Bhasker, C. R., Merrick, W. C. & Sen, G. C. Viral stress-inducible protein p56 inhibits translation by blocking the interaction of eIF3 with the ternary complex eIF2.GTP.Met-tRNAi. *J. Biol. Chem.* **278**, 39477–39482 (2003).
25. Coley, S. E. *et al.* Recombinant mouse hepatitis virus strain A59 from cloned, full-length cDNA replicates to high titers *in vitro* and is fully pathogenic *in vivo*. *J. Virol.* **79**, 3097–3106 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by National Institutes of Health grants U54 AI081680, U19 AI083019 and R01 AI074973 (to M.G. and M.S.D.), R01 AI56540 (to S.S.), U54 AI057160 (to R.M.B.) and U54 AI057158 (to P.Y.S.), R01 CA068782 (to G.C.S.) the Swiss National Science Foundation, 3100A0-118425/1, and the Novartis Foundation (to V.T. and R.Z.). We thank H. Virgin and B. Moss for reading the manuscript.

**Author Contributions** S.D., K.J.S., R.M.B., T.C.P., M.G., P.-Y.S. and M.S.D. designed the experiments. S.D., K.J.S., J.S., J.L., T.-Y.L. and H.D. performed the experiments. S.D., K.J.S., J.S., R.M.B., M.G., P.-Y.S. and M.S.D. analysed the data. S.Y., V.F., G.C.S., W.B.K., R.Z. and V.T. provided key reagents and expertise. S.D., K.J.S., R.M.B., T.C.P., M.G., P.-Y.S. and M.S.D. wrote and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.S.D. ([diamond@borcim.wustl.edu](mailto:diamond@borcim.wustl.edu)).

# Molecular coupling of *Tsix* regulation and pluripotency

Pablo Navarro<sup>1,2</sup>, Andrew Oldfield<sup>1,3</sup>, Julie Legoupi<sup>1</sup>, Nicola Festuccia<sup>2</sup>, Agnès Dubois<sup>1</sup>, Mikael Attia<sup>1</sup>, Jon Schoorlemmer<sup>4</sup>, Claire Rougeulle<sup>1,3</sup>, Ian Chambers<sup>2</sup> & Philip Avner<sup>1</sup>

The reprogramming of X-chromosome inactivation during the acquisition of pluripotency *in vivo* and *in vitro*<sup>1</sup> is accompanied by the repression of *Xist*<sup>2</sup>, the trigger of X-inactivation<sup>3</sup>, and the upregulation of its antisense counterpart *Tsix*<sup>4</sup>. We have shown that key factors supporting pluripotency—Nanog, Oct4 and Sox2—bind within *Xist* intron 1 in undifferentiated embryonic stem cells (ESC) to repress *Xist* transcription<sup>5</sup>. However, the relationship between transcription factors of the pluripotency network and *Tsix* regulation has remained unclear<sup>5,6</sup>. Here we show that *Tsix* upregulation in embryonic stem cells depends on the recruitment of the pluripotent marker Rex1, and of the reprogramming-associated factors Klf4 and c-Myc, by the *DXPas34* minisatellite associated with the *Tsix* promoter. Upon deletion of *DXPas34*, binding of the three factors is abrogated and the transcriptional machinery is no longer efficiently recruited to the *Tsix* promoter. Additional analyses including knock-down experiments further demonstrate that Rex1 is critically important for efficient transcription elongation of *Tsix*. Hence, distinct embryonic-stem-cell-specific complexes couple X-inactivation reprogramming and pluripotency, with Nanog, Oct4 and Sox2 repressing *Xist* to facilitate the reactivation of the inactive X, and Klf4, c-Myc and Rex1 activating *Tsix* to remodel *Xist* chromatin<sup>7–10</sup> and ensure random X-inactivation upon differentiation<sup>1</sup>. The holistic pattern of *Xist/Tsix* regulation by pluripotent factors that we have identified suggests a general direct governance of complex epigenetic processes by the machinery dedicated to pluripotency.

X-inactivation reprogramming in female mice is a model for the epigenetic processes underlying the acquisition of pluripotency<sup>1</sup>. The inactivation of the paternal X chromosome that characterizes the earliest cleavage-stages of development is followed by X-chromosome reactivation in the pluripotent inner cell mass of the blastocyst<sup>11,12</sup>. During differentiation, X-inactivation is established randomly on either the paternal or the maternal X chromosome<sup>1</sup>. The developmental plasticity of X-inactivation during early embryogenesis is paralleled by female induced pluripotent stem (iPS) cells<sup>13</sup>: the inactive X is reactivated in iPS cells and random X-inactivation *de novo* established upon loss of pluripotency<sup>14</sup>. Hence, two distinct processes affect X-inactivation during the generation of pluripotency: the reactivation of the inactive X *per se*, probably initiated by the repression of *Xist* by Nanog, Oct4, and Sox2 (refs 5,15), and the acquisition by both X chromosomes of an equal competence for future random X-inactivation, a process directly controlled by *Tsix*. Indeed, inactivation of *Tsix* in ESC<sup>8</sup> and embryos<sup>10</sup> leads to drastic, stable remodelling of *Xist* chromatin in *cis*<sup>7–10</sup>, associated with the systematic upregulation of *Xist* from the *Tsix*-null allele upon differentiation. Thus maximal *Tsix* activity is required in pluripotent cells to erase inherited *Xist* chromatin modifications and provide each X chromosome with equal probabilities of *Xist* upregulation during differentiation<sup>16</sup>.

It has been proposed that the high levels of *Tsix* transcription characterizing ESC depend on binding of Oct4 and Sox2 (ref. 6), but not

Nanog, at *DXPas34* (ref. 17) and at *Xite*<sup>18</sup> two enhancers of *Tsix*<sup>19</sup>. Oct4 and Sox2, by simultaneously controlling *Xist* and *Tsix*, could be acting at the top of the X-inactivation regulatory hierarchy in pluripotent cells. Uncertainties, however, surround this hypothesis because (1) the reported binding of Oct4 and Sox2 to *DXPas34*, the strong embryonic-stem-specific enhancer of *Tsix*, is not reproducible (Supplementary Fig. 1), (2) binding levels at *Xite* are very low (Supplementary Fig. 1) and (3) *Xite* is a weak enhancer of *Tsix*<sup>18</sup>. It appears likely that Oct4 and Sox2 play a minor role, if any, in the establishment of *Tsix* transcription in undifferentiated ESC. In agreement, *Tsix* remains unaffected after 24 h of Oct4 knockdown, whereas *Xist* upregulation is already established (Supplementary Fig. 1). Thus additional factors are implicated in controlling *Tsix* transcription in undifferentiated ESC. Three pluripotency factors, Klf4, c-Myc and Rex1, attracted our attention.

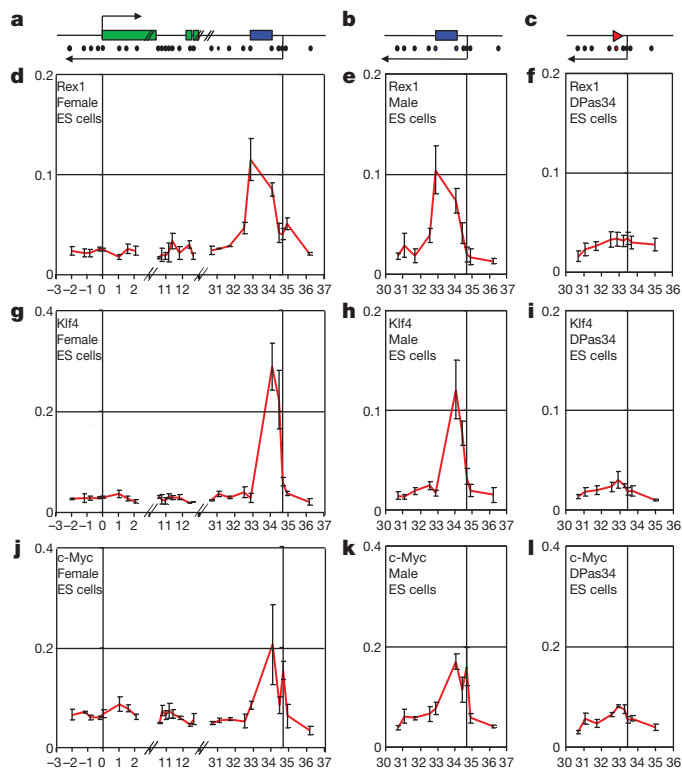
Because Oct4 and Sox2 directly mediate repression of *Xist* in ESC, we were interested to extend our analysis to Klf4 and c-Myc, the two other factors that are commonly used to generate iPS cells<sup>13</sup>. Rex1 is a marker of pluripotency whose deletion is associated with decreased *Tsix* expression, as revealed by microarray analysis<sup>20</sup>. During iPS cell generation, *Tsix* re-expression temporally correlates with the induction of *Rex1* (ref. 21).

We initially determined whether binding activity of Rex1, Klf4 and c-Myc could be detected at the *Xist/Tsix* region (Fig. 1a). We found binding to the *Tsix* 5' region for all three factors, in both female (Fig. 1 d, g, j) and male (Fig. 1 e, h, k) embryonic stem cells. Only Rex1 displays binding at both ends of *DXPas34*, which suggests that Rex1 is recruited within *DXPas34* itself whereas Klf4 and c-Myc are bound between *DXPas34* and the *Tsix* promoter. As expected given its chromatin immunoprecipitation (ChIP) profile, Rex1 binding is lost upon the targeted deletion of *DXPas34* in male embryonic stem cells ( $\Delta$ Pas34 cell line<sup>17</sup>; Fig. 1 c, f). Strikingly, Klf4 and c-Myc binding is similarly affected in  $\Delta$ Pas34 (Fig. 1 i, l), which suggests that *DXPas34* influences Klf4 and c-Myc recruitment. This drastic perturbation in transcription factor binding correlates with a 90% reduction in *Tsix* RNA levels<sup>17</sup>, mediated by a strong reduction of RNAPII recruitment at the *Tsix* promoter (Supplementary Fig. 2). Thus *DXPas34* orchestrates the recruitment of Rex1, Klf4 and c-Myc to activate *Tsix* transcription in ESC.

We next silenced *Oct4* in ZhbTc4.1 ESC to induce the loss of pluripotency<sup>22</sup> and the consequent downregulation of Rex1 and Klf4 (Fig. 2a). *Tsix* was downregulated (Fig. 2a) by transcriptional mechanisms (Fig. 2b, c), and binding of Rex1, Klf4 and c-Myc was reduced (Fig. 2d–f). Similar observations were made in terminally differentiated, *Tsix*-silenced<sup>9</sup>, mouse embryonic fibroblasts (Fig. 2g–i). Rex1, Klf4 and c-Myc are therefore critically important to couple *Tsix* regulation to pluripotency. We also analysed trophoblast stem cells, which display reduced *Tsix* transcription<sup>9</sup>, and found that only Rex1 is absent from the *Tsix* 5' region (Fig. 2g–i). Although the existence of differentiation-dependent repressors or additional embryonic-stem-specific activators cannot be excluded, this suggests that Rex1 plays

<sup>1</sup>Unité de Génétique Moléculaire Murine, URA 2578, Institut Pasteur, 75724 Paris Cedex 15, France. <sup>2</sup>Medical Research Council (MRC) Centre Development in Stem Cell Biology, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JQ, UK. <sup>3</sup>UMR 7216 Epigénétique et Destin Cellulaire, CNRS/Université Paris Diderot, Case 7042, 75205 Paris Cedex 13, France. <sup>4</sup>ARAID Foundation and Instituto Aragonés de Ciencias de la Salud, Departamento de Anatomía y Embriología, Facultad de Veterinaria, 50013 Zaragoza, Spain.

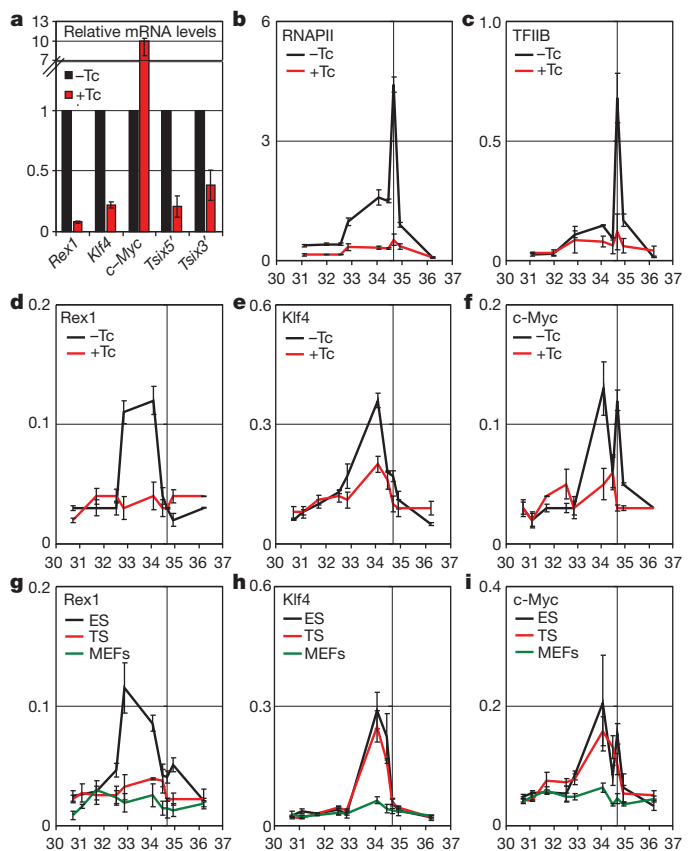




**Figure 1 | *DXPas34* orchestrates Rex1, Klf4 and c-Myc recruitment to the *Tsix* 5' region in pluripotent ESC.** **a–c**, Schematic representation of the subregions of the *Xist/Tsix* locus analysed by ChIP in female ESC (**a**), male ESC (**b**) and male  $\Delta$ Pas34 ESC (**c**). *Xist* exons are in green, the arrows indicate the direction of transcription of *Xist* (top) and *Tsix* (bottom). *DXPas34* is shown in blue. The location of each primer pair is indicated by a black circle. In **b**, the blue circles show the primer pairs flanking *DXPas34* that are absent in  $\Delta$ Pas34 ESC, whereas in **c** the red triangle shows the location of the loxP site remaining after *DXPas34* deletion, and the new primer pair designed at the position shown in red. ChIP analysis of Rex1 (**d–f**), Klf4 (**g–i**) and c-Myc (**j–l**) in the indicated cell lines. The x axis shows the genomic coordinates (in kilobases) relative to the *Xist* transcription start site. The vertical black bars represent the *Xist* and *Tsix* transcription start sites. The y axis shows the average percentage of immunoprecipitation. The number of biological replicates used was as follows: **d**,  $n = 2$ ; **e**,  $n = 3$ ; **f**,  $n = 6$ ; **g**,  $n = 3$ ; **h**,  $n = 3$ ; **i**,  $n = 6$ ; **j**,  $n = 3$ ; **k**,  $n = 3$ ; **l**,  $n = 6$ . All the results are expressed as means  $\pm$  s.e.m.

an important role in *Tsix* transcription in ESC. In agreement with this, transient *Rex1* interference demonstrates that *Tsix* is a rapid *Rex1*-responsive gene in ESC (Supplementary Fig. 3).

Next, we generated stable *Rex1* knockdown lines and confirmed that *Rex1* downregulation is accompanied by *Tsix* downregulation (Supplementary Fig. 3). The stably interfered clone displaying the highest level of *Rex1* knockdown, in which *Tsix* expression and *Rex1* binding are reduced by half (Fig. 3a, b), was analysed in detail. We observed that neither *Klf4* nor *c-Myc* are downregulated (Fig. 3a), nor affected in their binding to the *Tsix* 5' region (Fig. 3b). No drastic effect on RNAPII recruitment at the *Tsix* promoter (Fig. 3c), or on TFIIB, H3K4Me3 and H3K9Ac levels (Supplementary Fig. 3), was observed. Thus *Rex1* is not an essential factor for recruiting the transcriptional machinery at the *Tsix* promoter, or for triggering the accumulation of chromatin marks characteristic of transcription initiation<sup>23</sup>. Analysis of RNAPII (Fig. 3c) and H3K36Me3 (Fig. 3d and Supplementary Fig. 3) across the *Tsix* transcription unit did, however, reveal reduced levels in *Rex1* knockdown cells, in particular at the *Tsix* 3' end. Because the amount of 3'-end-associated RNAPII at the *Tsix* locus was previously proved to be a good readout of *Tsix* transcription<sup>7</sup>, and H3K36Me3 levels correlate well with transcriptional activity<sup>24</sup>, we conclude that *Rex1* is required for efficient elongation of *Tsix*. Statistical analysis of



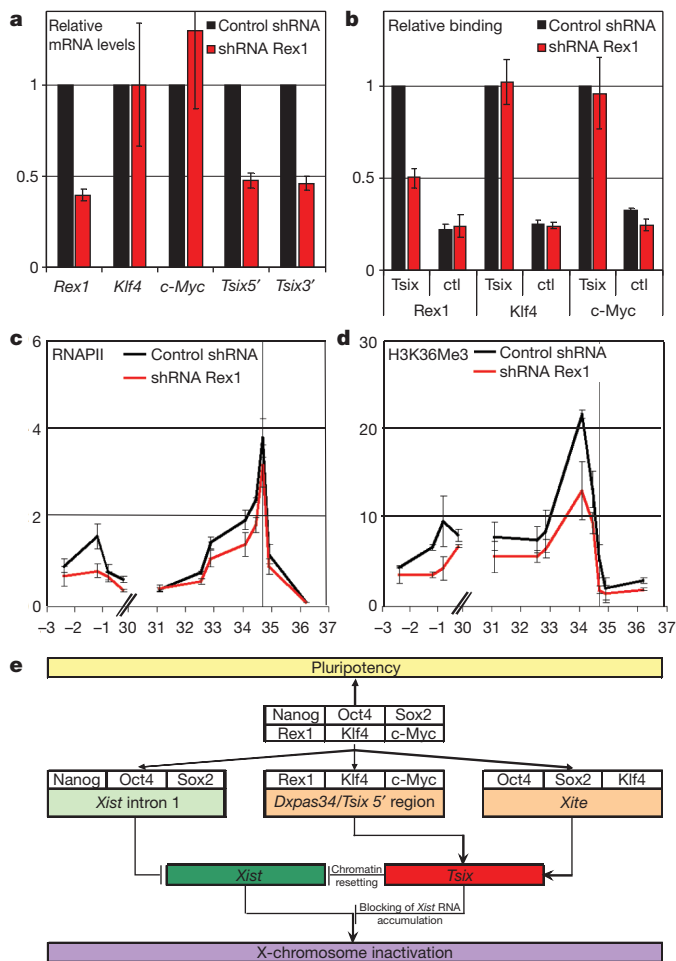
**Figure 2 | Developmentally induced loss of Rex1, Klf4 and c-Myc binding correlates with *Tsix* repression.** **a**, Relative RNA levels of the indicated genes in undifferentiated (–Tc, black, set to one) and differentiating (96 h Tc treatment, +Tc, red) ZhbTc4.1 cells ( $n = 4$ ). **b–f**, ChIP analysis of RNAPII (**b**), TFIIB (**c**), Rex1 (**d**), Klf4 (**e**) and c-Myc (**f**) across the *Tsix* 5' region in the same cellular conditions ( $n = 2$ ). **g–i**, Analysis of Rex1, Klf4 and c-Myc binding at the *Tsix* 5' region in ESC (black,  $n = 3$ ), trophoblast stem cells (F2, red,  $n = 3$ ), and mouse embryonic fibroblasts (green,  $n = 3$ ). The x axis shows the genomic coordinates (in kilobases) relative to the *Xist* transcription start site. All the results are expressed as means  $\pm$  s.e.m.

the ChIP profiles provides further support to the reduction of RNAPII and H3K36Me3 levels (Supplementary Fig. 3).

*c-Myc* has been shown to be a global regulator of transcription elongation in ESC (ref. 25), and affects in particular *Tsix* (Supplementary Fig. 4). This suggests that *Rex1* and *c-Myc* functionally interact to bring about efficient *Tsix* elongation. The fact that *c-Myc* remains associated to the *Tsix* 5' region in *Rex1* knockdown cells indicates that *Rex1* acts downstream of *c-Myc*. The differential segregation of *Rex1*/*c-Myc* from Nanog/Oct4/Sox2 targets<sup>26</sup> further indicates that *Rex1* might act as a global regulator of transcription elongation in ESC by providing developmental specificity to *c-Myc* function.

How these factors biochemically interact at the *Tsix* 5' region remains unknown. Given that (1) Yy1 (a factor evolutionary related to *Rex1* (ref. 27)) and Ctcf show positive binding at both extremities of *DXPas34*, (2) Sp1 binds at the *Tsix* 5' region and (3) interactions between these factors have been previously reported (Supplementary Fig. 5), we propose that *DXPas34* acts as a DNA platform directly recruiting *Rex1*, Ctcf and Yy1, which in turn facilitate the recruitment of Sp1, Klf4 and *c-Myc*. If *Rex1* and *c-Myc* regulate *Tsix* elongation, the parallel recruitment of the RNAPII and Klf4 at the *Tsix* 5' region observed in our experimental conditions leads us to speculate that Klf4 might be critically important to load the transcriptional machinery at the *Tsix* promoter.

Persistent binding of Nanog, Oct4 and Sox2 to *Xist* intron 1 in  $\Delta$ Pas34 ESC (Supplementary Fig. 6) maintains transcriptional *Xist*



**Figure 3 | Rex1 is required for efficient elongation of *Tsix* transcription.** **a**, Relative gene expression of a stable clone (clone 1a in Supplementary Fig. 3) expressing an shRNA against *Rex1* ( $n = 4$ ). **b**, ChIP analysis of *Rex1*, *Klf4* and *c-Myc* binding at the positions identified as providing maximal binding in Fig. 1 (*Tsix*) and at a negative control position (ctl) corresponding to Tpg; see Supplementary Fig. 8) in control (black) and *Rex1*-interfered cells (red,  $n = 3$ ). **c**, ChIP analysis of the RNAPII across the *Tsix* 3' and 5' regions in control and *Rex1*-interfered cells ( $n = 3$ ,  $m = 5$ ). **d**, ChIP analysis of H3K36Me3 across the *Tsix* 3' and 5' regions in control and *Rex1*-interfered cells ( $n = 2$ ,  $m = 3$ ). The  $x$  axis shows the genomic coordinates (in kilobases) relative to the *Xist* transcription start site. All the results are expressed as means  $\pm$  s.e.m. **e**, Transcriptional network coupling pluripotency regulators to X-inactivation. In a previous study we showed that Nanog, Oct4 and Sox2 bind *Xist* intron 1 to repress *Xist* transcription in ESC. Here we have shown that the pluripotency-associated *Rex1* protein, in conjunction with the reprogramming factors *Klf4* and *c-Myc*, binds to the *DXPas34/Tsix* 5' region to confer *Tsix* maximal activity in ESC. Moreover, Oct4, Sox2 and *Klf4* bind to *Xite*, a weak enhancer of *Tsix*. Whilst Nanog, Oct4 and Sox2 suppress *Xist* transcription, facilitating the reactivation of the inactive X, *Rex1*, *Klf4* and *c-Myc* transactivate *Tsix*, which, in turn, modifies the *Xist* chromatin structure to render all *Xist* alleles epigenetically indistinguishable and allow random *Xist* transcription upon differentiation. *Tsix* may additionally block *Xist* RNA accumulation at the post-transcriptional level of regulation. We conclude that the road to pluripotency and the path of X-inactivation regulation during both development and *in vitro* reprogramming experiments are directly coupled through the stringent control of the two main non-coding actors by distinct pluripotency-associated regulatory complexes.

silencing in the absence of *Tsix* transcription (Supplementary Fig. 2). Conversely, *Nanog*- and *Oct4*-inducible mutant ESC retain normal regulation of *Tsix*<sup>5</sup>. Moreover, *Xist* upregulation is observed from both wild-type and *Tsix*-deleted alleles upon *Oct4* knockdown (Supplementary Fig. 1). Based on these results, we propose that two distinct pluripotency-related complexes act independently but in

parallel to specify the reactivation of the inactive X and the resetting of the epigenetic conditions required for *de novo* random X-inactivation, through their direct regulation of *Xist* and *Tsix*, respectively (Fig. 3e). Although Oct4, Sox2 and *Klf4* show only low-level binding at *Xite* (Fig. 3e and Supplementary Fig. 5), inspection of available ChIP-Seq data sets<sup>28</sup> indicates that other genes located within the X-inactivation centre, including *Rnf12*<sup>29</sup>, are targets of the pluripotency-associated machinery. Further complexities of the molecular system coupling pluripotency and X-inactivation are therefore to be expected.

Interestingly, the connection of pluripotency regulators with *Xist* and *Tsix* may also apply to other epigenetic phenomena. Examination of ChIP-Seq data sets shows abundant binding of pluripotent factors at known imprinting centres including that of the *Dlk1-Dio3* cluster (Supplementary Fig. 7), which is inappropriately regulated during iPS cell generation<sup>30</sup>. Hence, over and above their importance for pluripotency and self-renewal, the pluripotent factors may be key components of other, more specific epigenetic processes occurring in pluripotent cells, notably in the germ line. Cohorts of pluripotency regulators could be involved, as we previously hypothesized for *Xist*<sup>1,5</sup>, in either the erasure and/or the establishment of epigenetic imprints, both at *Tsix* and at other imprinted loci.

## METHODS SUMMARY

Female ESC: LF2; male ESC: CK35;  $\Delta$ Pas34 ESC: #BH9 and #BD7; male trophectoderm stem cells: F2; male mouse embryonic fibroblasts: derived from embryonic day 13.5 embryos.

ChIP and PCR with reverse transcription were performed as previously described<sup>5</sup>. Chromatin and RNA preparations were isolated in parallel from the same culture batches.

Transient transfections of *Oct4* short interfering RNA (siRNA) (Dharmacon) and the *Rex1* short hairpin RNA (shRNA) expressing vectors were performed using a nucleofector (Lonza) and the manufacturer's protocol (program A30). Results were normalized to the RNA levels of cells nucleofected with non-targeting siRNA (Dharmacon), or with a vector expressing a shRNA against *Gfp*.

Stable integration of shRNA vectors was performed after electroporation of LF2 ESC, and selection in hygromycin B for 2 weeks. Resistant clones were individually expanded. Clones carrying the shRNA vectors against *Rex1* or *Gfp* were generated and analysed in parallel.

For primer sequences and antibody information see Supplementary Fig. 8.

In the figure legends, ' $n$ ' indicates the number of independent cell cultures analysed to control for biological variation. When indicated, ' $m$ ' shows the number of experiments performed with the ' $n$ ' independent extracts. All the results are expressed as means  $\pm$  s.e.m.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 June; accepted 6 September 2010.

- Navarro, P. & Avner, P. When X-inactivation meets pluripotency: an intimate rendezvous. *FEBS Lett* **583**, 1721–1727 (2009).
- Borsani, G. et al. Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325–329 (1991).
- Penny, G. D. et al. Requirement for *Xist* in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
- Lee, J. T., Davidow, L. S. & Warshawsky, D. *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. *Nature Genet.* **21**, 400–404 (1999).
- Navarro, P. et al. Molecular coupling of *Xist* regulation and pluripotency. *Science* **321**, 1693–1695 (2008).
- Donohoe, M. E. et al. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128–132 (2009).
- Navarro, P. et al. A role for non-coding *Tsix* transcription in partitioning chromatin domains within the mouse X-inactivation centre. *Epigenetics Chromatin* **2**, 8 (2009).
- Navarro, P., Page, D. R., Avner, P. & Rougeulle, C. *Tsix*-mediated epigenetic switch of a CTCF-flanked region of the *Xist* promoter determines the *Xist* transcription program. *Genes Dev.* **20**, 2787–2792 (2006).
- Navarro, P. et al. *Tsix* transcription across the *Xist* gene alters chromatin conformation without affecting *Xist* transcription: implications for X-chromosome inactivation. *Genes Dev.* **19**, 1474–1484 (2005).
- Sado, T., Hoki, Y. & Sasaki, H. *Tsix* silences *Xist* through modification of chromatin structure. *Dev. Cell* **9**, 159–165 (2005).
- Mak, W. et al. Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303**, 666–669 (2004).

12. Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 664–666 (2004).
13. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
14. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
15. Chambers, I. & Tomlinson, S. R. The transcriptional foundation of pluripotency. *Development* **136**, 2311–2322 (2009).
16. Navarro, P. & Avner, P. An embryonic story: analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells. *Bioessays* **32**, 581–588 (2010).
17. Vigneau, S. *et al.* An essential role for the DXPas34 tandem repeat and Tsix transcription in the counting process of X chromosome inactivation. *Proc. Natl Acad. Sci. USA* **103**, 7390–7395 (2006).
18. Ogawa, Y. & Lee, J. T. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol. Cell* **11**, 731–743 (2003).
19. Stavropoulos, N., Rowntree, R. K. & Lee, J. T. Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. *Mol. Cell Biol.* **25**, 2757–2769 (2005).
20. Masui, S. *et al.* Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Dev. Biol.* **8**, 45 (2008).
21. Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6**, e253 (2008).
22. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372–376 (2000).
23. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
24. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
25. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
26. Kim, J. *et al.* An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
27. Kim, J. D., Faulk, C. & Kim, J. Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Res.* **35**, 3442–3452 (2007).
28. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
29. Jonkers, I. *et al.* RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell* **139**, 999–1011 (2009).
30. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175–181 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Clerc for discussions. P.N. was initially supported by recurrent funding from the Institut Pasteur and then by the Royal Society (a Newton International Fellowship). P.A. was supported by recurrent funding from the Centre National de la Recherche Scientifique and the Institut Pasteur, contracts 05-JCJC-0166-01 and 07-BLAN-0047-01 from the Agence Nationale de la Recherche and funding from the EU Epigenome Network of Excellence. C.R. was supported by the INSERM 'Avenir' programme and by the European Research Council 'starting grant' programme. Research in I.C.'s laboratory was supported by the Wellcome Trust, by the EU Framework 7 project "EuroSyStem" and by a Medical Research Council studentship (to N.F.).

**Author Contributions** P.N. conceived the study, designed, performed and analysed the experiments, and co-wrote the manuscript. A.O., J.L., N.F., M.A. and A.D. provided technical help. C.R., I.C. and P.A. provided financial and conceptual support. P.A. conceived the study and co-wrote the manuscript. All authors read and approved the final manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to P.A. ([pavner@pasteur.fr](mailto:pavner@pasteur.fr)) or P.N. ([pablo.navarro@ed.ac.uk](mailto:pablo.navarro@ed.ac.uk)).



## METHODS

**RNA extraction and random reverse transcription.** Cells ( $1 \times 10^6$ – $5 \times 10^6$ ) were lysed in TRIzol (Invitrogen) and RNA was then extracted with chloroform and precipitated with isopropanol. After DNase treatment (Qiagen), RNA was re-extracted with phenol/chloroform, precipitated with ethanol, re-suspended in water and quantified.

RNA (1–4  $\mu$ g) was used per reverse transcription reaction. RNA was denatured in the presence of 1  $\mu$ g of random hexamers (Roche) for 5 min at 90 °C, and reverse transcribed in a final volume of 20  $\mu$ l with 100 U of SuperScriptII (Invitrogen) at 42 °C for 60 min followed by heat inactivation at 70 °C for 15 min. Synthesized complementary DNAs were diluted in 280  $\mu$ l of water and stored at –20 °C until used.

**Chromatin extraction and chromatin immunoprecipitation.** Twenty million cells were re-suspended in 3 ml of pre-warmed DMEM-FCS 10% and crosslinked with 1% formaldehyde (Sigma) for 10 min at room temperature. The reaction was quenched with 0.125 mM glycine for 5 min at room temperature. Cells were spun down for 3 min at 200g at 4 °C, and washed twice with cold PBS1X (Invitrogen). Cell pellets were then vigorously re-suspended in 300  $\mu$ l of Sweeling buffer (5 mM Pipes pH 8, 85 mM KCl) freshly supplemented with 1 $\times$  protease inhibitor cocktail (Roche) and 0.5% NP-40. The suspension was incubated for 20 min on ice with occasional gentle shaking. One microlitre of suspension was used to check for the completeness of total nuclei extraction under the microscope. Nuclei were spun down in 15-ml conical tubes for 10 min at 400g at 4 °C and re-suspended in 1.5 ml of TSE150 (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCl pH8, 150 mM NaCl) buffer, freshly supplemented with 1 $\times$  protease inhibitor cocktail. Samples were sonicated at 4 °C in 15-ml conical tubes using a Bioruptor (Diagenode) for five cycles of 10 min divided into 30 s on/30 s off subcycles at maximum power. Chromatin was then transferred into 1.5 ml tubes and centrifuged for 30 min at 15,340g at 4 °C. Soluble chromatin was divided into aliquots and stored at –80 °C until use. Twenty microlitres were used for quantity and quality controls of the DNA.

Twenty micrograms of DNA were used for each ChIP. For each experiment, the required amount of chromatin was defrosted (generally between 40 and 100  $\mu$ g – 1 to 5 ChIPs per sample) and pre-cleared for 1 h 30 min with rotation at 4 °C in 1 ml of TSE150 with 50  $\mu$ l of pA/pG sepharose beads (Sigma) 50% slurry, previously blocked with 500  $\mu$ g ml<sup>–1</sup> of molecular grade BSA (Roche) and 1  $\mu$ g ml<sup>–1</sup> of yeast tRNA (Invitrogen). Pre-cleared chromatin was transferred into fresh tubes after centrifugation for 1 min at 800g and divided into aliquots accordingly. Twenty micrograms of diluted chromatin were used for input DNA extraction and precipitation. Immunoprecipitation with specific antibodies (1–5  $\mu$ g each; see Supplementary Fig. 8) was performed overnight with rotation at 4 °C, in a final volume of 500  $\mu$ l. Immunocomplexes were recovered with 50  $\mu$ l of blocked pA/pG sepharose beads 50% slurry for 1 h 30 min with rotation at 4 °C. Beads were recovered by centrifugation for 1 min at 800g and washed at room temperature in 1 ml of TSE150, TSE500 (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCl pH8, 500 mM NaCl), washing buffer (10 mM Tris-HCl pH8, 0.25M LiCl, 0.5% NP40, 0.5% Na-Deoxycholate, 1 mM EDTA), and TE (10 mM Tris-HCl pH8, 1 mM EDTA). Each wash was performed for 5 min with rotation at room temperature. After the last wash, elution was performed in 100  $\mu$ l of elution buffer (1%

SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8) for 15 min at 65 °C after vigorous vortexing. Eluates were collected after centrifugation for 1 min at 15,340g, and the beads rinsed in 150  $\mu$ l of TE-SDS1%. After centrifugation for 1 min at 15,340g the supernatant was pooled with the corresponding first eluate. Crosslinking of ChIP and input fractions was reversed overnight at 65 °C, followed by proteinase K treatment (Invitrogen), extraction with phenol/chloroform and precipitation with ethanol. DNA pellets corresponding to the input fractions were re-suspended in 300  $\mu$ l of water, whereas those corresponding to the ChIP fraction were re-suspended in 100  $\mu$ l in the case of transcription factors, or in 300  $\mu$ l in the case of histone modifications.

**Transient siRNA knockdowns.** The medium of subconfluent embryonic stem cell cultures was changed 6 h before nucleofection. After cell collection, five million cells were pelleted in individual tubes, washed with PBS1X (Invitrogen) and re-suspended in 90  $\mu$ l of completed nucleofection solution (Amaxa). This cellular suspension was mixed with 10  $\mu$ l of siRNA (150 nM) or shRNA expressing vectors (4  $\mu$ g) and transferred into nucleofection cuvettes (Amaxa) that were placed in the nucleofection device. Program A30, which was used for all experiments, consistently gave more than 80% efficiency as evaluated by the nucleofection of a green fluorescent protein-expressing vector, with around 50% immediate mortality. Nucleofected cells were collected using the pipettes provided by Amaxa into 500  $\mu$ l of prewarmed embryonic stem cell medium, and transferred into 25-cm<sup>2</sup> gelatinized flasks containing 10 ml of prewarmed embryonic stem cell medium. Twenty-four hours later, cells were collected for analysis and RNA extraction.

**Rex1 and Gfp siRNAs.** The siRNAs (shRNA1 *Rex1*: 5'-ACGGATACC TAGAGTGCATCA, shRNA2 *Rex1*: 5'-CACGGAGAGCTCGAACTAAA, shRNA *Gfp*: 5'-AAGCGCGATCACATGGTCCTG) were designed using SiDE (<http://side.bioinfo.cnio.es>).

**Real-time PCR analysis.** Two systems of PCR analysis were exploited. All analyses except those corresponding to the *Rex1* knockdowns were analysed in 96-well plates using a StepOnePlus PCR machine (Applied Biosystems) and the Power Sybr Green PCR Master Mix (Applied Biosystems). *Rex1* knockdown experiments were analysed in 384-well plates with a 480 LightCycler (Roche) using LightCycler 480 SYBR Green I Master (Roche). All reactions were performed in duplicate. Five microlitres of DNA were used per reaction.

Standard curves of all primers were performed to check for efficient amplification (above 90%). Melting curves were also performed to verify production of single DNA species with each primer pair. All primer sequences are available in Supplementary Fig. 8.

Relative levels of expression in each assay were obtained through the  $\Delta\Delta C_t$  method, using (1) *ArpoP0* mRNA levels as a reporter in all experiments except in *Rex1* knockdowns, in which *Tbp* was used, and (2) the appropriate control cell line or cellular state as the reference sample.

Enrichment levels in ChIP assays are expressed as a percentage of immunoprecipitation relative to the input. Essentially, the  $\Delta C_t$  method was used to calculate a ChIP over input ratio that was corrected by the appropriate dilution factor of each analysed fraction, and multiplied by 100 to get the percentage of immunoprecipitation.

# Integrating carbon–halogen bond formation into medicinal plant metabolism

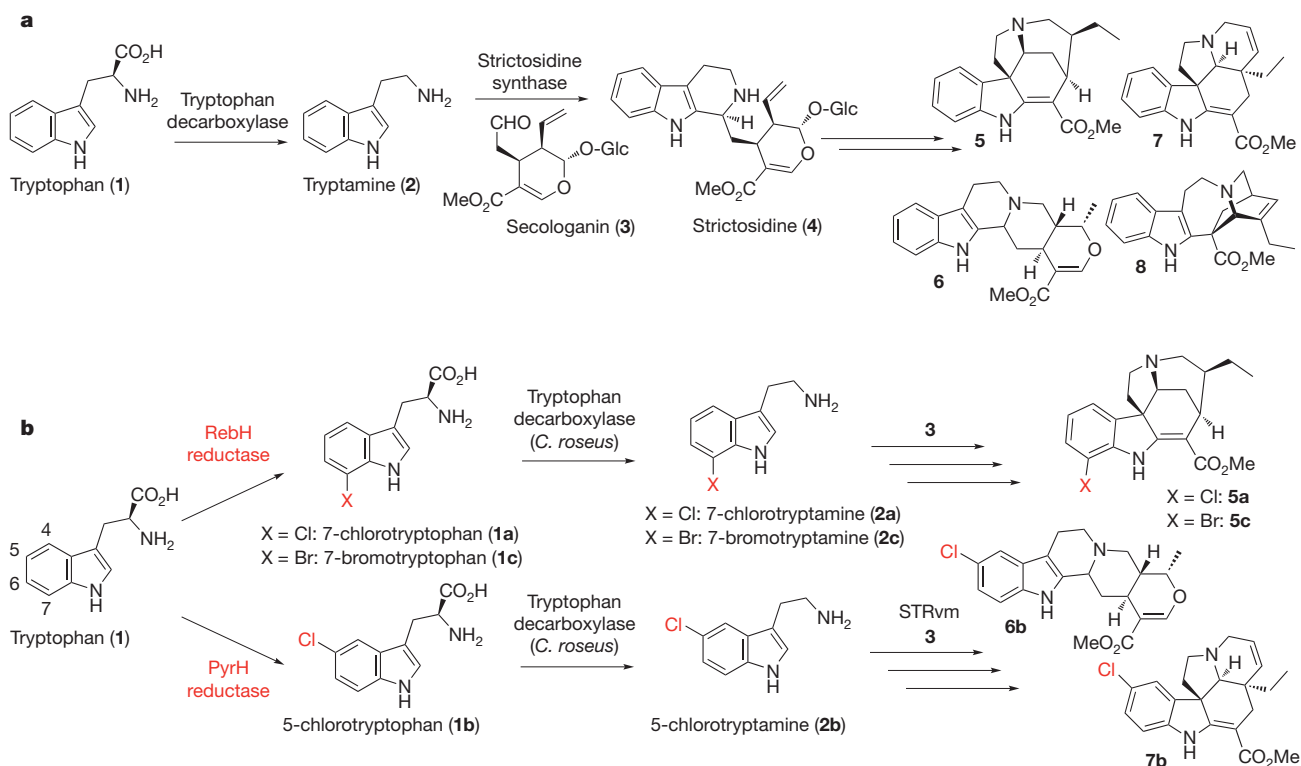
Weerawat Runguphan<sup>1\*</sup>, Xudong Qu<sup>1†\*</sup> & Sarah E. O'Connor<sup>1</sup>

Halogenation, which was once considered a rare occurrence in nature, has now been observed in many natural product biosynthetic pathways<sup>1</sup>. However, only a small fraction of halogenated compounds have been isolated from terrestrial plants<sup>2</sup>. Given the impact that halogenation can have on the biological activity of natural products<sup>1</sup>, we reasoned that the introduction of halides into medicinal plant metabolism would provide the opportunity to rationally bioengineer a broad variety of novel plant products with altered, and perhaps improved, pharmacological properties. Here we report that chlorination biosynthetic machinery from soil bacteria can be successfully introduced into the medicinal plant *Catharanthus roseus* (Madagascar periwinkle). These prokaryotic halogenases function within the context of the plant cell to generate chlorinated tryptophan, which is then shuttled into monoterpene indole alkaloid metabolism to yield chlorinated alkaloids. A new functional group—a halide—is thereby introduced into the complex metabolism of

*C. roseus*, and is incorporated in a predictable and regioselective manner onto the plant alkaloid products. Medicinal plants, despite their genetic and developmental complexity, therefore seem to be a viable platform for synthetic biology efforts.

Numerous halogenase enzymes from soil bacteria have been identified and characterized extensively<sup>1,3–5</sup>. Two of these flavoenzymes, PyrH<sup>6,7</sup> and RebH<sup>8–11</sup>, chlorinate the indole ring of tryptophan in the five and seven positions, respectively. Transferring these enzymes into other natural product pathways would allow site-specific incorporation of halogens onto a range of tryptophan-derived alkaloid products<sup>12</sup>, provided that the downstream enzymes could accommodate the chlorinated tryptophan precursor.

*Catharanthus roseus* produces a wide variety of monoterpene indole alkaloids<sup>13</sup> (Fig. 1a). This metabolic pathway begins with the conversion of tryptophan (1) to tryptamine (2) by tryptophan decarboxylase<sup>14</sup>. Tryptamine then condenses with the iridoid terpene



**Figure 1 | Monoterpene indole alkaloid biosynthesis.** **a**, Tryptophan (1) is decarboxylated by tryptophan decarboxylase to yield tryptamine (2), which reacts with secologanin (3) to form strictosidine (4). After numerous rearrangements, strictosidine (4) is converted into a variety of monoterpene indole alkaloids, such as 19,20-dihydroakuammicine (5), ajmalicine (6), tabersonine (7) and catharanthine (8). These compounds have a variety of

pharmacological activities<sup>24–26,30</sup>. Me, CH<sub>3</sub>; Glc, glucose. **b**, RebH and PyrH, along with a partner reductase, halogenate the indole ring of tryptophan to yield chlorotryptophan. Here we show that after transformation of these enzymes into *C. roseus*, the halogenated tryptophans **1a** and **1b** can be decarboxylated by tryptophan decarboxylase (*C. roseus*) to form the chlorotryptamines **2a** and **2b** and then converted into chlorinated monoterpene indole alkaloids.

<sup>1</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>†</sup>Present address: State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China.

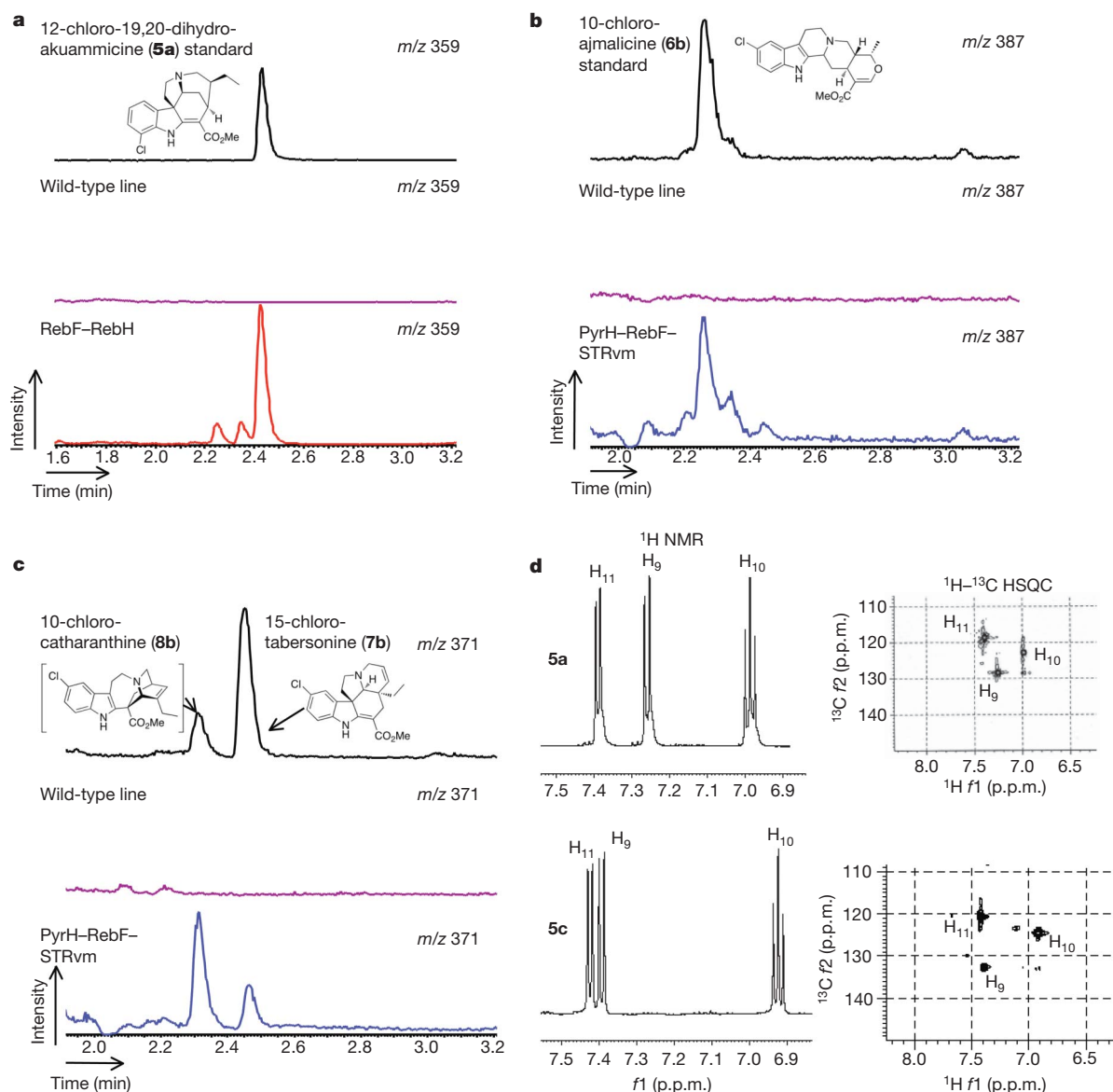
\*These authors contributed equally to this work.

secologanin (**3**) to form a biosynthetic intermediate strictosidine (**4**), which is subsequently functionalized in *C. roseus* to form over 100 alkaloids, including the anticancer agent vinblastine<sup>13</sup>. Previous work has shown that when *C. roseus* cell culture is supplemented with a variety of halogenated tryptamines, the corresponding halogenated alkaloid analogues are produced in isolable yields<sup>15,16</sup>. If prokaryotic halogenases could function in the eukaryotic plant cell, and if tryptophan decarboxylase could convert halogenated tryptophan into halogenated tryptamine, then *C. roseus* would produce chlorinated alkaloids *de novo* (Fig. 1b).

Because RebH and PyrH do not turn over tryptamine, this strategy requires that tryptophan decarboxylase from *C. roseus* recognize halogenated tryptophan. We assayed tryptophan decarboxylase from *C. roseus* *in vitro* with tryptophan ( $K_m = 51.7 \pm 9.2 \mu\text{M}$  (Michaelis constant),  $k_{\text{cat}} = 5.1 \pm 0.1 \text{ min}^{-1}$  (turnover number),  $k_{\text{cat}}/K_m =$

$0.099 \mu\text{M}^{-1} \text{ min}^{-1}$ ), 7-chlorotryptophan (**1a**;  $K_m = 499 \pm 74 \mu\text{M}$ ,  $k_{\text{cat}} = 1.6 \pm 0.04 \text{ min}^{-1}$ ,  $k_{\text{cat}}/K_m = 0.00327 \mu\text{M}^{-1} \text{ min}^{-1}$ ) and 5-chlorotryptophan (**1b**;  $K_m = 538 \pm 48 \mu\text{M}$ ,  $k_{\text{cat}} = 2.5 \pm 0.08 \text{ min}^{-1}$ ,  $k_{\text{cat}}/K_m = 0.00455 \mu\text{M}^{-1} \text{ min}^{-1}$ ) (Supplementary Figs 1 and 2). The activity of the enzyme suggested that halogenated tryptophan could be decarboxylated *in vivo*.

When considering how to merge the prokaryotic biosynthetic machinery with the plant alkaloid pathway, we chose to transfer the halogenase enzymes into *C. roseus* rather than move the plant biosynthetic enzymes into a microbial host. Most of the monoterpene indole alkaloid biosynthetic genes have not been identified, making heterologous expression of this pathway impossible at present. Moreover, we note that reconstitution of plant alkaloid pathways continues to be a challenge<sup>17,18</sup>. Many alkaloids use complex starting materials (such as secologanin) that are only produced by a few specialized plants, so



**Figure 2 | Chlorinated alkaloids in *C. roseus* hairy root culture.** **a**, LC-MS chromatograms showing 12-chloro-19,20-dihydroakuammicine (**5a**;  $m/z$  359) in RebF-RebH hairy roots (red trace), contrasted with control cultures transformed with no plasmid (purple trace). An authentic standard of **5a** validated the structural assignment (black trace; Supplementary Figs 30 and 31). **b**, Chromatograms showing 10-chloroajmalicine (**6b**) in RebF-PyrH-STRvm hairy roots (blue trace), contrasted with control cultures (purple trace). An authentic standard of **6b** is shown<sup>22</sup> (black trace). **c**, Chromatograms

showing 15-chlorotabersonine (**7b**) in RebF-PyrH-STRvm hairy roots (blue trace), contrasted with control cultures (purple trace). The other major peak at  $m/z$  371 had an exact mass and ultraviolet spectrum consistent with a chlorinated analogue of catharanthine<sup>22</sup> (**8**) (Supplementary Fig. 29). **d**,  $^1\text{H}$  NMR and  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear single quantum coherence (HSQC) spectra of **5a** and **5c**.  $f_1$  and  $f_2$ , chemical shifts in the  $^1\text{H}$  and  $^{13}\text{C}$  dimensions, respectively.



reconstitution of plant alkaloid pathways must also include biosynthesis of these precursors. For example, ajmalicine (**6**; Fig. 1a), one of the simplest of the monoterpene indole alkaloids, requires an estimated 14 discrete enzymes for biosynthesis from tryptophan and the terpene geraniol<sup>13</sup>; reconstitution of a pathway of this length is a significant engineering problem. Therefore, we believe that exploring approaches in the host plant is an important aspect of alkaloid metabolic engineering efforts.

To produce 7-chlorotryptophan *in planta*, we generated an expression construct containing codon-optimized complementary DNA encoding the 7-tryptophan chlorinase RebH and its required partner flavin reductase, RebF, in a plant expression vector (pCambia1300), both under the control of constitutive cauliflower mosaic virus (CaMV) 35S promoters. For production of 5-chlorotryptophan, an expression construct encoding the 5-chlorinating enzyme PyrH, along with RebF as the partner reductase, was generated. No signal sequence was added to the halogenase genes, to ensure that RebH, PyrH and RebF would produce chlorinated tryptophan in the cytosol, where it would most readily encounter the decarboxylase, which is also localized in the cytosol<sup>19</sup> (Supplementary Figs 3–5).

We used *Agrobacterium rhizogenes* to generate hairy root culture of *C. roseus* transformed with the halogenase genes<sup>20</sup>. One of the early biosynthetic enzymes, strictosidine synthase, cannot turn over 5-chlorotryptamine<sup>21</sup> (**2b**). Therefore, when transforming *C. roseus* with *pyrH* and *rebF*, we also introduced a mutant of strictosidine synthase (STRvm) that can convert 5-chlorotryptamine to 10-chlorostrictosidine<sup>16,22</sup> (**4b**). After a selection process, we cultivated the transformed root culture on standard Gamborg's B5 plant medium and monitored chlorinated alkaloids using liquid chromatography/mass spectrometry (LC–MS). We observed formation of chlorinated tryptophans **1a** and **1b** and chlorinated alkaloids in both the RebH–RebF and PyrH–RebF–STRvm hairy root lines (Fig. 2 and Supplementary Figs 6–15). These results indicate that RebH, PyrH and the partner reductase function productively in the plant cell environment, demonstrating that the flavin halogenases are highly transportable among kingdoms. Because chlorinated alkaloid production was observed in the transformed lines, we conclude that tryptophan decarboxylase can competently turn over halogenated tryptophan substrates *in vivo*.

Hairy roots transformed with RebH and RebF, which produce 7-chlorotryptophan, yielded a major chlorinated product at *m/z* 359 (Fig. 2a). An authentic standard of 12-chloro-19,20-dihydroakuammicine (**5a**) co-eluted with this compound. Natural products containing the akuammicine scaffold have a variety of pharmacological activities<sup>23–25</sup>. Although the parent compound, 19,20-dihydroakuammicine (**5**) has been isolated in good yields from other plants<sup>26</sup>, it is not a major alkaloid in *C. roseus* hairy root culture. However, when wild-type *C. roseus* cell lines were incubated with 7-chlorotryptamine, 12-chloro-19,20-dihydroakuammicine was also the major chlorinated product (Supplementary Fig. 16). Therefore, the predominance of 12-chloro-19,20-dihydroakuammicine in the RebH–RebF hairy root line is probably due to substrate specificity of downstream enzymes for 7-chlorotryptamine. A hairy root line transformed with the 5-chlorotryptophan enzyme system, PyrH, RebF and STRvm, produced a variety of chlorinated alkaloids (Fig. 2b–d). Two representative chlorinated alkaloids, 10-chloroajmalicine (**6b**) and 15-chlorotabersonine (**7b**), were identified by co-elution with authentic standards<sup>21</sup>.

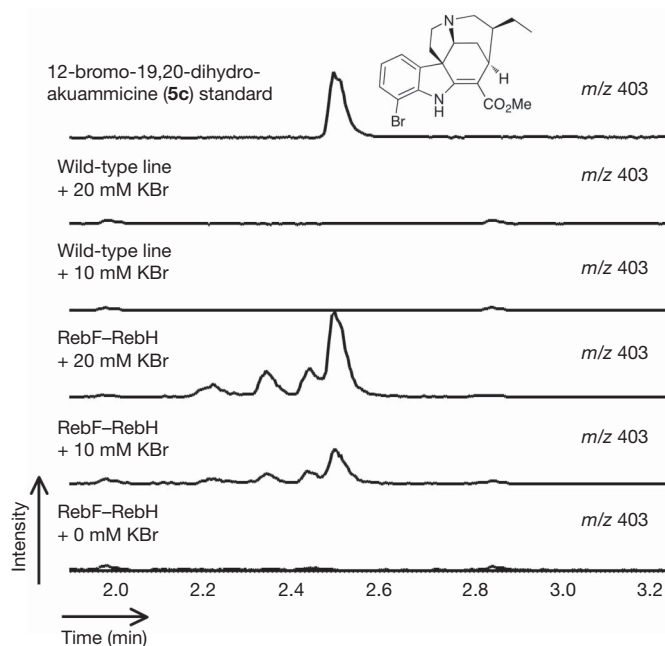
Chlorinated alkaloid production seemed to be stable over the course of at least six subcultures. The alkaloid 12-chloro-19,20-dihydroakuammicine was produced at  $26 \pm 4$   $\mu\text{g}$  per gram of fresh root weight of a representative cell line averaged over six subcultures. For comparison, wild-type cell lines produced  $\sim 25$   $\mu\text{g}$  per gram of fresh tissue weight of chlorinated alkaloids when the medium was supplemented with 200  $\mu\text{M}$  7-chlorotryptamine (**2a**). Similarly, 10-chloroajmalicine and 15-chlorotabersonine (**7b**) were produced at  $2.8 \pm 0.9$  and  $4.0 \pm 1.0$   $\mu\text{g}$  per gram of fresh root weight, respectively, for a representative cell line averaged over four subcultures (Supplementary Figs

12 and 14). Different concentrations of KCl (3  $\mu\text{M}$ –20 mM) were added to the medium, but increasing amounts of exogenous chloride salt did not significantly affect the yields of chlorinated alkaloids (Supplementary Figs 17 and 18).

Previous reports demonstrated that RebH can use bromide to yield brominated tryptophan<sup>8</sup> (**1c**). To assess the capacity of RebH for bromination *in vivo*, we supplemented a low-chloride cell culture medium with KBr. The *in vitro* halide specificity of RebH correlated with the products generated *in vivo*, as we observed the formation of a compound that co-eluted with an authentic standard of 12-bromo-19,20-dihydroakuammicine (**5c**) ( $21 \pm 8$  and  $49 \pm 20$   $\mu\text{g}$  per gram of fresh root weight with 10 mM and 20 mM KBr supplementation, respectively; Fig. 3). In contrast, supplementation of the medium with KI failed to yield either iodinated tryptophan or iodinated alkaloids. Again, this correlated with *in vitro* studies showing that RebH does not accept iodide as a substrate<sup>8</sup> (Supplementary Figs 19–22).

We also measured the transcript levels of the heterologous enzymes by real-time PCR with reverse transcription. The production of halogenated compounds depended on the expression of both RebF and RebH or PyrH. Notably, when the strictosidine synthase mutant STRvm was not expressed in the PyrH–RebF hairy root lines, we observed accumulation of 5-chlorotryptophan (representative cell line,  $9 \pm 1$   $\mu\text{g}$  per gram of fresh root weight) and 5-chlorotryptamine (representative cell line,  $20 \pm 9$   $\mu\text{g}$  per gram of fresh root weight), but did not observe downstream alkaloids (Supplementary Figs 23 and 24).

Tryptophan does not seem to accumulate in either wild-type or transformed hairy roots. However, accumulation of 7-chlorotryptophan ( $50 \pm 12$   $\mu\text{g}$  per gram of fresh root weight for a representative RebH–RebF cell line) and 5-chlorotryptophan ( $8 \pm 2$   $\mu\text{g}$  per gram of fresh root weight for a representative PyrH–RebF–STRvm cell line) was observed, suggesting that decarboxylation of chlorinated tryptophan is a bottleneck *in vivo*, a step that could potentially be subjected to future engineering efforts. This is consistent with the 30-fold-lower catalytic efficiency of the decarboxylase enzyme for halogenated



**Figure 3 | Extracted LC–MS chromatograms showing the presence of 12-bromo-19,20-dihydroakuammicine (**5c**; *m/z* 403) in RebF–RebH hairy roots.** Hairy roots are grown in medium supplemented with KBr (0–20 mM final concentration) for two weeks before alkaloid extractions. 12-bromo-19,20-dihydroakuammicine is not observed in control cultures transformed with no plasmid after incubation in KBr-supplemented medium. An authentic standard of 12-bromo-19,20-dihydroakuammicine is used to validate the structural assignment (Supplementary Figs 29, 30 and 32).

tryptophan *in vitro*. The morphologies of the halogen producing lines were thicker and slower growing than those of wild-type lines (Supplementary Fig. 25). Because tryptophan serves as the precursor for other small-molecule metabolites, we speculate that chlorinated tryptophan may be diverted into other pathways such as auxins. Notably, 4-chloro indole acetic acid, which is found in several species of pea, has altered activity relative to the auxin indole acetic acid<sup>27,28</sup>.

Medicinal plants produce a wide range of complex natural products but generate relatively few halogenated compounds; chlorinated or brominated compounds are not found among the approximately 3,000 known monoterpene indole alkaloids produced by plants in the Apocynaceae, Rubiaceae and Loganiaceae families. The halogenation of natural products often has profound effects on the bioactivity of the compound, and can serve as a useful handle for further chemical derivatization<sup>1,29</sup>. Despite the metabolic and developmental complexity of plant tissue, transformation of these prokaryotic genes led to the regioselective incorporation of halides into the alkaloid products of the existing plant pathway. Notably, the yield of chlorinated alkaloids in the most productive lines (~26 µg per gram of fresh weight of plant tissue) is only 15-fold lower than the yield of total natural alkaloids (compounds 5, 6, 7 and 8) from wild-type tissue (~420 µg per gram of fresh weight of plant tissue) (Supplementary Fig. 26). The ease with which we engineered the successful production of chlorinated alkaloids in *C. roseus*, a plant with limited genetic characterization, indicates that medicinal plants can provide a viable platform for synthetic biology.

## METHODS SUMMARY

Structural characterization is shown in Supplementary Figs 27–32 and Supplementary Tables 1 and 2.

**Generation of transgenic *C. roseus* hairy root cultures.** We transformed the expression construct pCAMRebHRebF into *A. rhizogenes* ATCC 15834 by electroporation (1-mm cuvette, 1.25 kV), and we co-transformed pCAMPyrHRebF and pCAMSTRvm into *A. rhizogenes* ATCC 15834 by electroporation. Transformation of *C. roseus* seedlings with the generated *Agrobacterium* strains was performed as previously reported<sup>20</sup>.

**Evaluation of alkaloid production in transgenic *C. roseus* hairy roots.** Every transgenic hairy root line that survived hygromycin selection medium was evaluated for alkaloid production. Transformed hairy roots were grown in Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins, 30 g l<sup>-1</sup> sucrose, 6 g l<sup>-1</sup> agar, pH 5.7). The total chloride concentration in Gamborg's B5 formulation was ~1 mM. We ground three-week-old hairy roots with a mortar, pestle and 106-µm acid-washed glass beads in methanol (10 ml g<sup>-1</sup> fresh weight of hairy roots). The crude natural product mixtures were filtered through 0.2-mm cellulose acetate membrane (VWR) and subsequently subjected to LC-MS analysis. Hairy roots transformed with wild-type *A. rhizogenes* lacking the plasmid were also evaluated.

**Brominated alkaloid production in transgenic *C. roseus* hairy roots.** We grew a selected transformed hairy root line for two weeks in low-chloride solid medium (67 mg l<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 353 mg l<sup>-1</sup> Ca(NO<sub>3</sub>)<sub>2</sub>·4H<sub>2</sub>O, 61 mg l<sup>-1</sup> MgSO<sub>4</sub>, 1,250 mg l<sup>-1</sup> KNO<sub>3</sub>, half-strength Murashige and Skoog micronutrient salts and full-strength Murashige and Skoog vitamins, 3 µM total chloride concentration). Hairy roots were transferred to the same medium supplemented with either potassium bromide or potassium iodide (10–20 mM final concentration) and cultivated for an additional two weeks. They were then processed and alkaloid production was analysed as described above (Supplementary Figs 12–15). We performed experiments in duplicate.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 April; accepted 20 September 2010.

Published online 3 November 2010.

1. Neumann, C. S., Fujimori, D. G. & Walsh, C. T. Halogenation strategies in natural products biosynthesis. *Chem. Biol.* **15**, 99–109 (2008).
2. Gribble, G. W. The diversity of naturally produced organohalogenes. *Chemosphere* **52**, 289–297 (2003).
3. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodiová, S. & Walsh, C. T. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem. Rev.* **106**, 3364–3378 (2006).

4. van Pée, K. H. & Patallo, E. P. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl. Microbiol. Biotechnol.* **70**, 631–641 (2006).
5. Blasiak, L. C. & Drennan, C. L. Structural perspective on enzymatic halogenation. *Acc. Chem. Res.* **42**, 147–155 (2009).
6. Zehner, S. *et al.* A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* **12**, 445–452 (2005).
7. Zhu, X. *et al.* Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J. Mol. Biol.* **391**, 74–85 (2009).
8. Yeh, E., Garneau, S. & Walsh, C. T. Robust *in vitro* activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl Acad. Sci. USA* **102**, 3960–3965 (2005).
9. Yeh, E. *et al.* Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry* **45**, 7904–7912 (2006).
10. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **46**, 1284–1292 (2007).
11. Bitto, E. *et al.* The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins* **70**, 289–293 (2008).
12. Sánchez, C. *et al.* Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc. Natl Acad. Sci. USA* **102**, 461–466 (2005).
13. O'Connor, S. E. & Maresh, J. Chemistry and biology of terpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23**, 532–547 (2006).
14. De Luca, V., Marineau, C. & Brisson, N. Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases. *Proc. Natl Acad. Sci. USA* **86**, 2582–2586 (1989).
15. McCoy, E. & O'Connor, S. E. Directed biosynthesis of alkaloid analogs in the medicinal plant *Catharanthus roseus*. *J. Am. Chem. Soc.* **128**, 14276–14277 (2006).
16. Bernhardt, P., McCoy, E. & O'Connor, S. E. Rapid identification of enzyme variants for reengineered alkaloid biosynthesis in periwinkle. *Chem. Biol.* **14**, 888–897 (2007).
17. Hawkins, K. M. & Smolke, C. D. Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. *Nature Chem. Biol.* **4**, 564–573 (2008).
18. Minami, H. *et al.* Microbial production of plant benzylisoquinoline alkaloids. *Proc. Natl Acad. Sci. USA* **105**, 7393–7398 (2008).
19. De Luca, V. & Cutler, A. J. Subcellular localization of enzymes involved in indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol.* **85**, 1099–1102 (1987).
20. Hughes, E. H., Hong, S.-B., Shanks, J. V., San, K.-Y. & Gibson, S. I. Characterization of an inducible promoter system in *Catharanthus roseus* hairy roots. *Biotechnol. Prog.* **18**, 1183–1186 (2002).
21. Rungtaphan, W. & O'Connor, S. E. Metabolic reprogramming of periwinkle plant culture. *Nature Chem. Biol.* **5**, 151–153 (2009).
22. Loris, E. A. *et al.* Structure-based engineering of strictosidine synthase: auxiliary for alkaloid libraries. *Chem. Biol.* **14**, 979–985 (2007).
23. Menzies, J. R., Paterson, S. J., Duwieuja, M. & Corbett, A. D. Opioid activity of alkaloids extracted from *Picralima nitida* (fam. Apocynaceae). *Eur. J. Pharmacol.* **350**, 101–108 (1998).
24. Frédérich, M. *et al.* Antiplasmodial activity of alkaloids from various *strychnos* species. *J. Nat. Prod.* **65**, 1381–1386 (2002).
25. Zhu, W.-M. *et al.* Components of stem barks of *Winchia calophylla* A. DC. and their bronchodilator activities. *J. Integr. Plant Biol.* **47**, 892–896 (2005).
26. Amat, M., Linares, A. & Bosch, J. A new synthetic entry to pentacyclic *Strychnos* alkaloids. Total synthesis of (+, -)-tubifolidine, (+, -)-tubifoline, and (+, -)-19,20-dihydroakumammicine. *J. Org. Chem.* **55**, 6299–6312 (1990).
27. Gandar, J. C. & Nitsch, C. Isolement de l'ester méthylique d'un acide chloro-3-indolylacétique à partir de graines immatures de pois, *Pisum sativum* L. *C. R. Acad. Sci. (Paris) Ser. D* **265**, 1795–1798 (1967).
28. Marumo, S., Hattori, H., Abe, H. & Munakata, K. Isolation of 4-chloroindolyl-3-acetic acid from immature seeds of *Pisum sativum*. *Nature* **219**, 959–960 (1968).
29. Deb Roy, A., Grischow, S., Cairns, N. & Goss, R. J. Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J. Am. Chem. Soc.* **132**, 12243–12245 (2010).
30. Li, S. *et al.* Assessment of the therapeutic activity of a combination of almitrine and raubasine on functional rehabilitation following ischaemic stroke. *Curr. Med. Res. Opin.* **20**, 409–415 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge support from the NIH (GM074820) and the American Cancer Society (RSG-07-025-01-CDD). We thank H.-Y. Lee and M. Tjandra for assistance with NMR characterizations and L. Li for high-resolution mass spectroscopy analysis.

**Author Contributions** All authors contributed to experimental design and data analysis. X.Q. initiated the project and its design, and performed steady-state kinetics. W.R. developed and implemented the transformation strategy and performed steady-state kinetics and metabolite analysis. All authors contributed to the preparation of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.E.O.C. (soc@mit.edu).



## METHODS

**Heterologous expression and purification of *C. roseus* tryptophan decarboxylase.** The tryptophan decarboxylase (TDC) gene (accession number M25151.1) was obtained by reverse-transcription PCR (RT-PCR) amplification of mRNA isolated from *C. roseus* hairy root culture (Invitrogen, Dynabeads mRNA direct kit) with PCR primers that introduce sites for NdeI and XhoI (underlined): 5'-AAAAAATATATGGGAGCATTGATTCACAA-3' and 5'-AAAAAATCTCGA GTCAAGCTTCTTTGAGCAAATC-3'. The PCR fragment was subcloned into the pGEM-T Easy Vector (Promega), and then excised and ligated into the NdeI/XhoI site of the pET28a-plasmid (Novagen). The resulting pET28a-TDC construct was subsequently transformed into BL21 (DE3) pLysS electrocompetent *Escherichia coli* (Promega). A single *E. coli* colony harbouring pET28a-TDC was inoculated in 5 ml lysogeny broth medium supplemented with kanamycin (0.05 mg l<sup>-1</sup>) and incubated overnight at 37 °C with shaking at 225 r.p.m. An aliquot of the overnight culture (1 ml) was then used to inoculate 100 ml lysogeny broth medium supplemented with kanamycin (0.05 mg l<sup>-1</sup>) and incubated at 37 °C with shaking at 225 r.p.m. until OD<sub>600</sub> 0.6 was reached. Cells were induced for overexpression by the addition of isopropyl-β-D-galactopyranoside (IPTG; final concentration, 1 mM) and the culture was allowed to continue growth for 16 h at 18 °C. Cells were harvested by centrifugation and lysed by sonication. The hexahistidine-tagged TDC was purified using Ni-NTA Spin Kit (Qiagen) using manufacturer's protocols (Supplementary Fig. 1). Eluted enzyme was subsequently buffer-exchanged into phosphate buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 100 mM NaCl, pH 8.0) and immediately assayed for activity. This enzyme was not stable after extended storage.

**Determining the steady-state kinetic constants of TDC for tryptophan substrate analogues, 5- and 7-chlorotryptophan (1b and 1a).** Steady-state kinetic constants of TDC for 5- and 7-chlorotryptophan (1b and 1a) (Amatek) were determined in phosphate buffer (0.1 M NaH<sub>2</sub>PO<sub>4</sub>, 3.5 mM β-mercaptoethanol, pH 8.5) containing 1 mM pyridoxal-5'-phosphate at 30 °C (0.3-ml reaction volume) with TDC concentrations appropriate for obtaining the initial rate of the reaction (0.6–0.9 μM). Aliquots (25 μl) were quenched in 1 ml methanol, containing yohimbine (500 nM) as an internal standard, at appropriate time points. The samples were centrifuged (13,000 r.p.m. (16,000g), 5 min) to remove particulates and then analysed by LC-MS. Samples were ionized by ESI with a Micromass LCT Premier TOF Mass Spectrometer. The liquid chromatography was performed on an Acquity Ultra Performance BEH C18, 1.7 μm, 2.1 × 100 mm column on a gradient of 10–90% acetonitrile/water (0.1% formic acid) over 5 min at a flow rate of 0.6 ml min<sup>-1</sup>. The appearance of the corresponding tryptamine analogues (either 5- or 7-chlorotryptamine) was monitored by peak integration and normalized to the internal standard. 5-chlorotryptamine was obtained from a commercial source (Alfa Aesar). 7-chlorotryptamine was synthesized as previously reported<sup>3</sup>. Eight substrate concentrations (200–2,500 μM) were tested for 7-chlorotryptophan substrate, and six substrate concentrations (200–1,200 μM) were tested for 5-chlorotryptophan substrate. Each concentration was assayed three times and the average values are reported with standard deviations. The data were fitted using nonlinear regression to the Michaelis–Menten equation using ORIGINPRO 7 (OriginLab). For reference, the kinetic constants for the natural substrate tryptophan (1) were also measured at concentrations ranging from 15 to 350 μM (Supplementary Fig. 2).

**Construction of halogenase plant expression vectors PyrH–RebF–STRvm and RebH–RebF.** The construction of plant expression vectors is summarized below and in Supplementary Fig. 3.

(i) The CaMV35S:Gus:NosPolyA fragment was obtained by PCR amplification of pCAMBIA1305.1 (Cambia) with forward and reverse PCR primers CaMV35S-NosPolyA that introduce sites for XbaI and KpnI at the 5' end, and PstI and SpeI at the 3' end (underlined): 5'-ACTTCTAGAGGTACCGGATCCTCTAGAGTCG ACCTGCAG-3' and 5'-ATTCTGCAGACTAGTCCCGATCTAGTAACATAG ATGACACCG-3'.

(ii) The tryptophan 5-halogenase gene (*pyrH*; accession number AAU95674) was obtained by PCR amplification of genomic DNA isolated from *Streptomyces rugosporus* NRRL 21084 with forward and reverse PCR primers CrPyrH that introduce sites for XhoI and NcoI at the 5' end, and SpeI and BstEII at the 3' end (underlined): 5'-ACTCTCGAGCCATGGATATCCGATCTGTGGTATCG-3' and 5'-ACTACTAGTGGTAACCTCATTGGATGCTGGCGAGGTA-3'.

(iii) The flavin reductase gene (*rebF*; accession number BAC15756) was obtained by PCR amplification of genomic DNA isolated from *Lechevalieria aerocolonigenes* ATCC 39243 with forward and reverse PCR primers CrRebF that introduce sites for XhoI and NcoI at the 5' end, and SpeI and PmlI at the 3' end (underlined): 5'-ACTCTCGAGCCATGGATACGATCGAGTTCGACAGAC-3' and 5'-ACT ACTAGTACCGTGTATCCCTCCGGTGTCCACAC-3'.

(iv) The tryptophan 7-halogenase gene (*rebH*; accession number BAC15758) was obtained by PCR amplification of genomic DNA isolated from *Lechevalieria*

*aerocolonigenes* ATCC 39243 with forward and reverse PCR primers CrRebH that introduce sites for SpeI and NcoI at the 5' end, and SpeI and PmlI at the 3' end (underlined): 5'-AAGACTACTAGTCCATGGATTCCGGCAAGATTGAC-3' and 5'-ACTACTAGTACCGTGTACGCGCCGTGCTGTTGCC-3'.

CaMV35S:Gus:NosPolyA, PyrH, RebH and RebF PCR fragments were individually ligated into pGEM-Teasy vector (Promega) to yield pGEMCaMV35S, pGEMPyrH, pGEMRebH and pGEMRebF, respectively.

(v) Codon-optimization for *C. roseus* was performed to ensure efficient expression of the prokaryotic halogenase and flavin reductase genes in plant cell culture. Using codon usage database software (<http://www.kazusa.or.jp/codon/>), the following codons were identified as occurring at low frequency (triplet, frequency per thousand): GCG, 4.8; CGG, 3.3; ACG, 4.0; TCG, 6.1; CCG, 6.0; and CCC, 6.7. Site-directed mutagenesis was performed using a Stratagene QuikChange Site-Directed Mutagenesis kit to replace rare codons with the most frequently occurring codons encoding the corresponding amino acids. Only codons that appeared within the first 300 nucleotides of the genes were subjected to mutagenesis.

The site-directed-mutagenesis primers (name, sequence) were as follows (the sites of mutation are underlined): PyrH-SDM-for1, 5'-GTGGGTGGTGGC ACTGCTGGCTGGATGACC-3'; PyrH-SDM-rev1, 5'-GGTCATCCAGCCAGC AGTGCCACCAACCCAC-3'; PyrH-SDM-for2, 5'-GACATGCGGCCGTATAC TACTGCTACCGCGATGAGCGCCGGC-3'; PyrH-SDM-rev2, 5'-GCCGGCG CTCATCGCGGTAGCAGTAGTGTACGCCGCATGTC-3'; RebH-SDM-for, 5'-CCCCAATCTGCAGACTGCTTCTTCGACTTCCTCGGA-3'; RebH-SDM-rev, 5'-TCCGAGGAAGTCAAGAAAGCAGTCTGCAGATTGGGG-3'; RebF-SDM-for1, 5'-ACCGCGCGCGATCAAGGGCTCTGATGAGCCTGTTTCCC-3'; RebF-SDM-rev1, 5'-GGGAAACAGGGCTCATCAGAGCCCTGTGATCGGCC CGGGT-3'; RebF-SDM-for2, 5'-CTCGTCTGCCTGAACAGGGCTAGCGGAA CGTTGCAC-3'; RebF-SDM-rev2, 5'-GTGCAACGTTCCGCTAGCCCTGTTC AGGCAGACGAG-3'.

(vi) pGEMCaMV35S was digested and ligated into the KpnI/EcoRI sites of pSP72 vector (Promega) to yield pSPCaMV35S.

(vii) pGEMRebH and pGEMRebF were digested and ligated into the NcoI/PmlI sites of pSPCaMV35S to yield pSPRebH and pSPRebF, respectively. Similarly, pGEMPyrH was digested and ligated into the NcoI/BstEII sites of pSPCaMV35S to yield pSPPyH. pSPRebF was then digested and ligated into the PstI site of pSP72 to yield pSPRebF-2.

(viii) pSPRebH and pSPPyH were digested and ligated into the XbaI/EcoRI sites of pSPRebF-2 to yield pSPRebHRebF and pSPPyHRebF, respectively. Finally, both pSPRebHRebF and pSPPyHRebF were digested and ligated into the SpeI site of pCAMBIA1300A to yield pCAMRebHRebF and pCAMPyrHRebF. pCAMBIA1300A was constructed by introducing an SpeI restriction into pCAMBIA1300 (Cambia). The site-directed-mutagenesis primers are (SpeI site underlined) 5'-CCCGCCTTCAGTTTAACTAGTCAGTGTGACAGGAT-3' and 5'-atcgtgtaaacactgACTAGTttaaactgaagcgagg-3'.

pCAMSTRvm was constructed as previously described<sup>21</sup>.

**Generation of transgenic *C. roseus* hairy root cultures.** The plant expression construct pCAMRebHRebF was transformed into *A. rhizogenes* ATCC 15834 by means of electroporation (1-mm cuvette, 1.25 kV). pCAMPyrHRebF and pCAMSTRvm were co-transformed into *A. rhizogenes* ATCC 15834 via electroporation (1-mm cuvette, 1.25 kV). Transformation of *C. roseus* seedlings with the generated *Agrobacterium* strains was performed as previously reported<sup>21</sup>. Briefly, 180–250 *C. roseus* seedlings (Vince Little Bright Eyes, Nature Hills Nursery) were germinated aseptically on Gamborg's B5 medium (full-strength basal salts, full-strength vitamins, 30 g l<sup>-1</sup> sucrose, pH 5.7) and grown in a 16-h light, 8-h dark cycle at 26 °C for 3 weeks. Seedlings were then wounded with extra-fine forceps at the stem tip, and 3–5 μl *A. rhizogenes* from a freshly grown liquid culture were inoculated on the wound.

Hairy roots appeared at the wound site 2–3 weeks after infection for about 80% of the seedlings infected. After hairy roots reached 1–4 cm in length (usually about 6 weeks after infection), they were excised and transferred to Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins, 30 g l<sup>-1</sup> sucrose, 6 g l<sup>-1</sup> agar, pH 5.7) containing hygromycin (0.03 mg ml<sup>-1</sup>) for selection and the antibiotic cefotaxime (0.25 mg ml<sup>-1</sup>) to remove remaining bacteria. The total chloride concentration in Gamborg's B5 formulation was 1 mM. All cultures were grown in the dark at 26 °C. After the hygromycin selection process, hairy roots were maintained in solid medium lacking both hygromycin and cefotaxime.

To adapt the line to liquid culture, approximately 200 mg of hairy roots (typically five 3–4-cm-long stem tips) from each line that grew successfully on solid medium were transferred to 50 ml of half-strength Gamborg's B5 liquid medium. The cultures were grown at 26 °C in the dark at 125 r.p.m. Hairy root growth in liquid medium seemed to be slower than that in solid medium. Hairy root transformants were screened for survival in solid medium supplemented with hygromycin. The number of transformants decreased significantly after solid medium



selection for each of the constructs transformed. Every line that grew in the selection medium was analysed for alkaloid production.

**Hairy root selection and adaptation processes.** For the plasmid pCAMRebH/RebF, the number of transformed hairy roots was 200 and the number of hairy roots after solid medium selection was 31. For the plasmid pCAMPyrH/RebF/STRvm, the number of transformed hairy roots was 140 and the number of hairy roots after solid medium selection was 57.

**Verification of transferred DNA integration by genomic DNA analysis.** To verify the integration of transferred DNA (T-DNA) into the plant genome, the genomic DNA from transformed hairy roots was isolated (Qiagen DNeasy kit) and then subjected to PCR amplification using T-DNA-specific primers with TDC primers serving as a positive control (see below). Specifically, for the pCAMRebH/RebF hairy roots, primers for PCR amplification were designed to amplify the complete TDC gene (TDC\_for and TDC\_rev), a 660-base-pair (bp) region of the RebH gene (RebH\_for and RebH\_rev), a 680-bp region of the RebF gene (RebF\_for and RebF\_rev) and an 800-bp region of the selection marker HPT gene (HPT\_for and HPT\_rev) (see below).

PCR primers for verification of T-DNA integration of transformed hairy roots were as follows (name, sequence): TDC\_for, 5'-AAAAACATATGGGCAGCATTGATCAACA-3'; TDC\_rev, 5'-AAAAAACTCGAGTCAAGCTTCTTTGAGCAATC-3'; RebH\_for, 5'-GTCCTCGATGCCGACCTCTTC-3'; RebH\_rev, 5'-GTACATGTCGATCTTCTCCTGC-3'; RebF\_for, 5'-TAGAGGACCTAACAGAAC-3'; RebF\_rev, 5'-CGTGACACTGGTCAGGGA-3'; HPT\_for, 5'-GCCTGA ACTCACCGCGACGTC-3'; HPT\_rev, 5'-CCTCCAGAAGAAGATGTTGGC-3'.

PCR amplification of genomic DNA from all of the selected transformed lines (pCAMRebH/RebF cell line 4, lanes 1–4; pCAMRebH/RebF cell line 5, lanes 5–8; pCAMRebH/RebF cell line 6, lanes 9–12; pCAMRebH/RebF cell line 10, lanes 13–16; and pCAMRebH/RebF cell line 11, lanes 17–20) was successful for all four sets of primers (Supplementary Fig. 4). PCR amplification of hairy root transformed with *A. rhizogenes* lacking the pCAMBIA vector (provided by Professor Jacqueline Shanks (Iowa State University) and Professor Carolyn Lee-Parsons (Northeastern University)) genomic DNA was successful only when TDC-specific primers were used (lanes 21–24). These results indicated that *rebH* and *rebF* were successfully incorporated into the *C. roseus* genome in all chosen lines.

For the pCAMPyrH/RebF/STRvm hairy roots, primers for PCR amplification were designed to amplify the complete TDC gene (TDC\_for and TDC\_rev), the complete PyrH gene (PyrH\_for and PyrH\_rev), a 680-bp region of the RebF gene (RebF\_for and RebF\_rev), a 440-bp region of the STRvm gene and an 800-bp region of the selection marker HPT gene (HPT\_for and HPT\_rev) (see below and Supplementary Fig. 5).

PCR primers for verification of T-DNA integration of transformed hairy roots were as follows (name, sequence): TDC\_for, 5'-AAAAACATATGGGCAGCATTTGATCAACA-3'; TDC\_rev, 5'-AAAAAACTCGAGTCAAGCTTCTTTGAGCAATC-3'; PyrH\_for, 5'-ATGATCCGATCTGTGGTG-3'; PyrH\_rev, 5'-TCATTGGATGCTGGCGAG-3'; RebF\_for, 5'-TAGAGGACCTAACAGAAC-3'; RebF\_rev, 5'-CGTGACACTGGTCAGGGA-3'; STRvm\_for, 5'-CCTTATTATTGAAAGAGCTACATATG-3'; STRvm\_rev, 5'-GCTAGAAACATAAGAAATTTCCCTTG-3'; HPT\_for, 5'-GCCTGA ACTCACCGCGACGTC-3'; HPT\_rev, 5'-CCTCCAGAAGAAGATGTTGGC-3'.

PCR amplification of genomic DNA from three of four of the selected transformed lines (pCAMPyrH/RebF/STRvm cell line 1, lanes 1–5; pCAMPyrH/RebF/STRvm cell line 3, lanes 6–10; pCAMPyrH/RebF/STRvm cell line 6, lanes 11–15) was successful for all five sets of primers (Supplementary Fig. 5). PCR amplification of genomic DNA from pCAMPyrH/RebF/STRvm cell line 7 (lanes 16–20) was successful when TDC, PyrH, RebF and HPT primers were used but not when STRvm primers were used. PCR amplification of hairy root transformed with *A. rhizogenes* lacking the pCAMBIA vector genomic DNA was successful only when TDC specific primers were used (lanes 21–25).

**Evaluation of alkaloid production in transgenic *C. roseus* hairy roots.** Every transgenic hairy root line that survived hygromycin selection medium was evaluated for alkaloid production. Transformed hairy roots were grown in Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins, 30 g l<sup>-1</sup> sucrose, 6 g l<sup>-1</sup> agar, pH 5.7). The total chloride concentration in Gamborg's B5 formulation was ~1 mM. Three-week-old hairy roots were ground with a mortar, pestle and 106 µm acid-washed glass beads in methanol (10 ml g<sup>-1</sup> of fresh hairy root weight). The crude natural product mixtures were filtered through 0.2-µm cellulose acetate membrane (VWR) and subsequently subjected to LC–MS analysis. Additionally, hairy roots transformed with wild-type *A. rhizogenes* lacking the plasmid were also evaluated.

These crude alkaloid mixtures were diluted 30:830 with methanol for mass spectral analysis. Samples were ionized by ESI with a Micromass LCT Premier TOF Mass Spectrometer. The liquid chromatography was performed on an Acquity Ultra Performance BEH C18, 1.7 µm, 2.1 × 100 mm column on a gradient

of 10–90% acetonitrile/water (0.1% TFA) over 5 min at a flow rate of 0.6 ml min<sup>-1</sup>. The capillary and sample cone voltages were 1,300 and 60 V, respectively. The desolvation and source temperature were 300 and 100 °C, respectively. The cone and desolvation gas flow rates were 60 and 800 l per hour, respectively. Analysis was performed with MASSLYNX 4.1. Accurate mass measurements were obtained in W-mode. The spectra were processed using the MASSLYNX 4.1 mass measure, in which the mass spectrum of peaks of interest was smoothed and centred with TOF mass correction, locking on the reference infusion of reserpine. Data for RebF–RebH lines are shown in extracted LC–MS chromatograms in Supplementary Figs 6–10.

**Feeding of 7-chlorotryptamine (2a) in control *C. roseus* hairy root cultures transformed with no plasmid.** Alkaloid accumulation levels in hairy roots transformed with RebH and RebF were compared with alkaloid accumulation levels in control hairy root fed with 7-chlorotryptamine. The control hairy root line was grown for 2 weeks in half-strength Gamborg's B5 solid medium. Hairy roots were then transferred to the same medium supplemented with 7-chlorotryptamine (2a; 0, 25, 50, 100, 200 and 750 µM final concentrations) and grown for a further 1 week. Hairy roots were then processed and alkaloid production analysed as described in the previous subsection (Supplementary Fig. 16). Feeding studies were performed in duplicate.

**Brominated alkaloid production in transgenic *C. roseus* hairy roots.** A selected transformed hairy root line was grown for 2 weeks in low-chloride solid medium (67 mg l<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 353 mg l<sup>-1</sup> Ca(NO<sub>3</sub>)<sub>2</sub>·4H<sub>2</sub>O, 61 mg l<sup>-1</sup> MgSO<sub>4</sub>, 1,250 mg l<sup>-1</sup> KNO<sub>3</sub>, half-strength Murashige and Skoog micronutrient salts and full-strength Murashige and Skoog vitamins, 3 µM total chloride concentration). Hairy roots were transferred to the same medium supplemented with either potassium bromide or potassium iodide (10–20 mM final concentration) and cultivated for an additional 2 weeks. Hairy roots were then processed and alkaloid production analysed as described in the previous subsection but one. Hairy roots transformed with wild-type *A. rhizogenes* lacking the plasmid were also evaluated (Supplementary Figs 19–22). Experiments were performed in duplicate.

**Purification and isolation of alkaloids from transformed TDC suppressed hairy roots supplemented with 7-chlorotryptamine and 7-bromotryptamine.** To obtain chlorinated and brominated alkaloid standards, root tips (10–15) from TDC suppressed hairy roots<sup>32</sup> were subcultured in six 50 ml Gamborg's B5 liquid medium (half-strength basal salts, full-strength vitamins, 30 g l<sup>-1</sup> sucrose, pH 5.7) and grown at 26 °C in the dark at 125 r.p.m. for 3 weeks before the medium was supplemented with either 7-chlorotryptamine (2a) or 7-bromotryptamine (2c) (750 µM final concentration). Both tryptamine analogue substrates were synthesized as previously reported<sup>33</sup>. After 2 weeks of co-cultivation, hairy roots were extracted as described above in methanol (10 ml g<sup>-1</sup> fresh hairy root weight). Alkaloid extracts were filtered, concentrated under vacuum and redissolved in 25% acetonitrile/water (0.1% TFA) (1 ml g<sup>-1</sup> of fresh hairy root weight).

For cultures supplemented with 7-chlorotryptamine (2a), the redissolved mixture was purified on a 10 × 20 mm Vydec reverse-phase column using a gradient of 25–52% acetonitrile/water (0.1% TFA) over 24 min. Alkaloids were monitored at 228 nm and fractions containing the alkaloid analogues of interest, as determined by the characteristic isotopic distribution expected for chlorinated molecules (<sup>35</sup>Cl/<sup>37</sup>Cl) from LC–MS analysis, were combined and concentrated under vacuum (Supplementary Fig. 27).

For cultures supplemented with 7-bromotryptamine (2c), similar procedures were performed to isolate alkaloids from transgenic hairy roots, except that the liquid chromatography method was extended to 26 min. Alkaloids were monitored at 228 nm and fractions containing the alkaloid analogues of interest, as determined by LC–MS analysis, were combined and concentrated under vacuum (Supplementary Fig. 28).

Isolated alkaloids from both feedings were analysed by LC–MS (same parameters as above), analytical high-performance liquid chromatography and, where possible, high resolution LC–MS (Supplementary Table 1), ultraviolet–visible spectroscopy (Supplementary Fig. 29), tandem MS–MS (Supplementary Fig. 30) and <sup>1</sup>H NMR, <sup>13</sup>C NMR and <sup>1</sup>H–<sup>13</sup>C HSQC using a Bruker AVANCE-600 NMR spectrometer equipped with a 5-mm 1H[13C,31P] cryoprobe (Supplementary Figs 31 and 32). Halogenated alkaloids generally displayed longer retention times than the natural alkaloids.

**Quantification of chlorinated alkaloid production in transformed hairy roots.** 12-chloro-19,20-dihydroakummicine (5a) and 12-bromo-19,20-dihydroakummicine (5c) standard curves were constructed by quantifying the peak areas of several concentrations (20–1,400 nM) of each alkaloid authentic standard using MASSLYNX 4.1. Similarly, 10-chloroajmalicine (6b), 15-chlorotabersonine (7b) and 10-chlorocatharanthine (8b) standard curves were constructed by quantifying the peak areas of several concentrations (20–1,400 nM) of each natural (that is, non-chlorinated) alkaloid authentic standard using MASSLYNX 4.1.

**Dependence of chlorinated alkaloid production on concentrations of sodium chloride.** A selected transformed hairy root line was grown for 2 weeks in low-chloride solid medium ( $67 \text{ mg l}^{-1}$   $(\text{NH}_4)_2\text{SO}_4$ ,  $353 \text{ mg l}^{-1}$   $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ ,  $61 \text{ mg l}^{-1}$   $\text{MgSO}_4$ ,  $1,250 \text{ mg l}^{-1}$   $\text{KNO}_3$ , half-strength Murashige and Skoog micronutrient salts and full-strength Murashige and Skoog vitamins,  $3 \mu\text{M}$  total chloride concentration). Hairy roots were transferred to the same medium supplemented with potassium chloride ( $0\text{--}20 \text{ mM}$  final concentration) and grown for a further 2 weeks. Hairy roots were then processed and alkaloid production was analysed as previously described (Supplementary Figs 17 and 18).

**Assessment of the stability of chlorinated alkaloid production in subsequent subcultures.** Ten root tips from hairy roots transformed with pCAMRebH/RebF and pCAMPyrH/RebF/STRvm were subcultured every 3 weeks in Gamborg's B5 solid medium (half-strength basal salts, full-strength vitamins,  $30 \text{ g l}^{-1}$  sucrose,  $6 \text{ g l}^{-1}$  agar, pH 5.7), and grown at  $26^\circ\text{C}$  in the dark. Alkaloids were isolated from 21-day-old hairy roots and analysed as described above (Supplementary Figs 11–14).

**Verification of expression of RebH, RebF and PyrH, STRvm enzymes by real-time RT-PCR.** Real-time RT-PCR was used to assess the expression levels of RebH and RebF. Expression levels in hairy roots infected with *A. rhizogenes* lacking the pCAMRebH/RebF construct were compared with expression levels in hairy roots harbouring pCAMRebH/RebF. Messenger RNA from transformed hairy roots was isolated and purified from contaminant DNA using a Qiagen RNeasy Plant Mini Kit and Rnase-free DnaseI, respectively. The resulting mRNA was then reverse-transcribed to cDNA using a Qiagen QuantiTect Reverse transcription kit and then subjected to PCR with specific primers (see below), a Qiagen SYBR Green PCR kit and a Biorad DNA Engine Opticon 2 system. The threshold cycle ( $C_T$ ) was determined as the cycle with a signal higher than that of the background plus 10 s.d. *Catharanthus roseus* 40S ribosomal protein S9 (Rps9), encoded by a house-keeping gene, was used to adjust the amount of the total mRNA in all samples. Real-time RT-PCR was performed in triplicate and the data are pictured as the relative expression levels of *rebH* and *rebF* mRNA in transgenic hairy roots as well as hairy roots lacking the pCAMBIA plasmid (Supplementary Fig. 23).

PCR primers for real-time RT-PCR of pCAMRebH/RebF transformed hairy roots were designed using the GenScript web tool (<http://www.genscript.com/ssl-bin/app/primer>), and were as follows (name, sequence): RebH\_for, 5'-GACGG

GCATCTACTTCGTCT-3'; RebH\_rev, 5'-TCGAACATCGTCTCGATCTC-3' (amplicon size, 117); RebF\_for, 5'-CTGATGAGCCTGTTCCCA-3'; RebF\_rev, 5'-CGTGACACTGGTCAGGGA-3' (amplicon size, 99); Rbps9\_for, 5'-TTGAGC CGTATCAGAAATGC-3'; Rbps9\_rev, 5'-CCCTCATCAAGCAGACCATA-3' (amplicon size, 122).

Real time RT-PCR was used to assess the expression levels of PyrH, RebF and STRvm. Expression levels in hairy roots infected with *A. rhizogenes* lacking the pCAMPyrH/RebF/STRvm construct were compared with expression levels in hairy roots harbouring pCAMPyrH/RebF/STRvm (Supplementary Fig. 24).

Chlorinated alkaloid production in a line lacking STRvm expression is shown in Supplementary Fig. 15. Photographs of the transformed roots are shown in Supplementary Fig. 25.

PCR primers for real-time RT-PCR of pCAMPyrH/RebF/STRvm transformed hairy roots were designed using GenScript web tool, and were as follows (name, sequence): PyrH\_for, 5'-GCCTGCTCATCAACCAGAC-3'; PyrH\_rev, 5'-CATC GCGGTAGCAGTAGTGT-3' (amplicon size, 137); RebF\_for, 5'-CTGATGAG CCTGTTTCCCA-3'; RebF\_rev, 5'-CGTGACACTGGTCAGGGA-3' (amplicon size, 99); STRvm\_for, 5'-TATTATTGAAAGAGCTACATATG-3'; STRvm\_rev, 5'-CTCTGCACTGCCTTTCTTG-3' (amplicon size, 134); Rbps9\_for, 5'-TTGA GCCGTATCAGAAATGC-3'; Rbps9\_rev, 5'-CCCTCATCAAGCAGACCATA-3' (amplicon size, 122).

**Quantification of natural alkaloids in wild-type roots.** The levels of natural alkaloids in wild type hairy roots was quantified as described in section on quantification of chlorinated alkaloid production in transformed hairy roots.

The levels of the four most abundant alkaloids found in these hairy roots, ajmalicine (6), tabersonine (7), catharanthine (8), as well as tryptophan (1) and tryptamine (2), were measured (Supplementary Fig. 26).

31. Hughes, E. H. *et al.* Characterization of an inducible promoter system in *Catharanthus roseus* hairy roots. *Biotechnol. Prog.* **18**, 1183–1186 (2002).
32. Runguphan, W. *et al.* Silencing of tryptamine biosynthesis for production of nonnatural alkaloids in plant culture *Proc. Natl Acad. Sci. USA* **106**, 13673–13678 (2009).
33. Schumacher, R. W. *et al.* Synthesis of didemnomolines A-D, N9-substituted  $\beta$ -carboline alkaloids from the marine ascidian *Didemnum* sp. *Tetrahedron* **55**, 935–942 (1999).

# WORLD WIRE WEB

*Information at your fingertips.*

BY GARETH D JONES

There was a time when Boyd could have had all the information he desired at his fingertips. It became impossible to keep track of the sources of information that threatened to overwhelm everyday life. That was before the Pulse permanently wiped out all electronics worldwide. Only those like Boyd who were septuagenarians at least, could remember those times of information overload. With his memory failing intermittently, a sign of weakness that he couldn't afford, Boyd missed those days immensely.

He snapped his fingers to get the attention of Carmichael, a man of great size, low ambition and unquestioning loyalty. The burly man strode across the dimly lit lounge and leaned over Boyd's chair. His bald head reflected the yellowish light of a wall lamp.

"Get down to the Web office." Boyd's voice was still strong, but croaked perpetually. "I need a name."

Carmichael carefully pulled a black notebook and a gilt pen from the inner pocket of his suit. He stood patiently, pen in hand.

"Fifty years ago. Ran the dog track across the docks. Had a boy with a missing finger. His girlfriend — tall, brunette. I need to know what she was called. See if you can track her down."

Carmichael made brief notes in his book, then closed it and returned it to his pocket. As Carmichael left the room, Boyd shuffled a collection of papers on the oval table before him and prepared to meet his lieutenants.

The door into the Web office swung open slowly, accompanied by a tinkling bell, and Celia looked up from the counter at the imposing figure who entered. The chubby smile slid from her face. Smart suit, impassive face, shorn head — he worked for the Guv'nor, that was obvious. The man moved with a grace surprising for his size and pulled out a small notebook as he approached the counter. He tore out a page and placed it on the wooden surface, turned it to face Celia and pushed it across to her.

"The Guv'nor needs this information." There was no threat, no intimidation, but Celia knew that she should ask no questions.

She took a moment to

**▶ NATURE.COM**  
Discuss this story  
online at:  
[go.nature.com/gbenem](http://go.nature.com/gbenem)

read the neatly printed words, then turned to her workstation. Alongside a chunky black typewriter was a brass lever that protruded through a slot in the desk. A total of 12 notches adjoining the slot were neatly labelled with possible destinations for the telegram. She engaged the lever and moved it up into the slot labelled 'Council Offices'. Pulleys and ratchets connected her teletyper to the telegraph line that led from her roof to the Council Offices several miles away.

She began typing her message with the word 'Urgent'. Everyone wanted their message to be dealt with urgently, but operators

room and stared out of the window. Celia turned back to her equipment, and willed the reply to arrive speedily. While she waited, she indulged in her regular daydream of working at one of the Web Hubs, where up to 100 destinations could be selected by a series of levers.

The reply came in 20 minutes, with information from the Land Registry and the Register Office, but with requests destined for elsewhere. Celia typed in the new destination and imagined her words racing along the lines to the North London Hub, and from there northward to the Lincoln Hub, then locally to the Lincolnshire Register Office.

By the time the request returned to Celia's machine it incorporated the name of the son, born in Lincolnshire; a note from a North London hospital of a fatal stabbing almost 50 years past matching the name — and a 'No Comment' from the Metropolitan Police.

The final comment was from an archivist at a North London newspaper. He had unearthed a report on the funeral, a picture of a small group of mourners, among them a tall brunette — a name.

Celia pulled the paper from the teletyper, circled the name and handed it over to Carmichael closer to two hours from when he had entered. He inclined his head politely and left without a word.

The meeting had gone well, although Boyd saw a predatory gleam in the eyes of some of his lieutenants. Carmichael arrived back as he was sipping the remains of a cup of tea. He took the paper and stared at the name. Annabelle. Yes, how could he have forgotten?

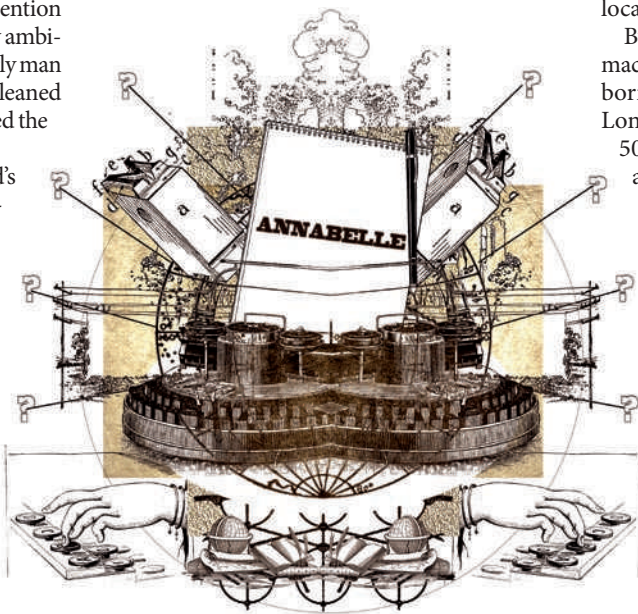
He pushed aside an assortment of papers and opened a large leather-bound notebook. A paragraph had been abandoned halfway down the page, and here Boyd took up the pen to continue his autobiography.

'Her name was Annabelle. When she ran off with the son of a dog-track owner, it started a feud that shook the whole borough. Her friend ran a boutique...'

"Carmichael, I need a name!"

Carmichael reached into his suit and once more pulled out his pen and notebook. ■

**Gareth D. Jones** is an environmental scientist who also writes stories and drinks lots of tea. His stories have appeared in 40 publications.



knew to use it only when absolutely necessary. With each key stroke, not only were her words typed on the carbonated paper, but the plunger attached to each key strummed the wire strung beneath. The plungers were marked with a series of grooves that represented the Morse code for that letter. The wires, one for each row, vibrated in time to the code. They were linked to a delicately balanced connector that danced to the rhythm and tapped out the message across the telegraph wire.

When the message was complete she pulled the paper from the machine and gave a copy to the unmoving Carmichael.

"How long?"

"Within the hour, usually."

"I'll wait."

Celia had been afraid he would say that. He took a seat in the corner of the waiting



# Quantitative reactivity profiling predicts functional cysteines in proteomes

Eranthie Weerapana<sup>1,2\*</sup>, Chu Wang<sup>1,2\*</sup>, Gabriel M. Simon<sup>1,2</sup>, Florian Richter<sup>3,4</sup>, Sagar Khare<sup>3,5</sup>, Myles B. D. Dillon<sup>2</sup>, Daniel A. Bachovchin<sup>1,2</sup>, Kerri Mowen<sup>2</sup>, David Baker<sup>3,4,5</sup> & Benjamin F. Cravatt<sup>1,2</sup>

Cysteine is the most intrinsically nucleophilic amino acid in proteins, where its reactivity is tuned to perform diverse biochemical functions. The absence of a consensus sequence that defines functional cysteines in proteins has hindered their discovery and characterization. Here we describe a proteomics method to profile quantitatively the intrinsic reactivity of cysteine residues en masse directly in native biological systems. Hyper-reactivity was a rare feature among cysteines and it was found to specify a wide range of activities, including nucleophilic and reductive catalysis and sites of oxidative modification. Hyper-reactive cysteines were identified in several proteins of uncharacterized function, including a residue conserved across eukaryotic phylogeny that we show is required for yeast viability and is involved in iron-sulphur protein biogenesis. We also demonstrate that quantitative reactivity profiling can form the basis for screening and functional assignment of cysteines in computationally designed proteins, where it discriminated catalytically active from inactive cysteine hydrolase designs.

Large-scale scientific endeavours such as genome sequencing and structural genomics are providing a wealth of new information on the full complement of proteins present in eukaryotic and prokaryotic organisms. Many of these proteins, however, remain partly or completely unannotated with respect to their biochemical activities<sup>1</sup>. New methods are therefore needed to characterize protein function on a global scale. Much effort is currently devoted to the characterization of post-translational modification events because these covalent adducts can have profound and dynamic effects on protein activity<sup>2</sup>. Another frequently overlooked parameter that defines functional 'hotspots' in the proteome is amino acid side-chain reactivity, which can vary by several orders of magnitude for a given residue depending on local protein microenvironment. Methods to measure side-chain reactivity en masse directly in complex biological systems have not yet been described, and as such, the reactive landscape of the proteome remains largely unexplored.

Among the protein-coding amino acids, cysteine is unique owing to its intrinsically high nucleophilicity and sensitivity to oxidative modification. The  $pK_a$  of the free cysteine thiol is between 8 and 9, meaning that only slight perturbations in the local protein microenvironment can result in ionized thiolate groups with enhanced reactivity at physiological pH<sup>3</sup>. Diverse families of enzymes use cysteine-dependent chemical transformations, including proteases, oxidoreductases and acyltransferases<sup>4</sup>. In addition to its role in catalysis, cysteine is subject to several forms of oxidative post-translational modification, including sulphenation (SOH), sulphination (SO<sub>2</sub>H), nitrosylation (SNO), disulphide formation and glutathionylation, which endow it with the ability to serve as a regulatory switch on proteins that is responsive to the cellular redox state<sup>5</sup>.

Functional cysteines, regardless of whether they are catalytic residues or sites of post-translational modification, do not conform to a canonical sequence motif, which complicates their systematic identification and characterization.  $pK_a$  measurements can identify cysteine residues with heightened nucleophilicity (or 'hyper-reactive' cysteines<sup>6,7</sup>), but

this requires purified protein and detailed kinetic and mutagenic experiments<sup>7,8</sup> that cannot be performed on a proteome-wide scale. Additional methods have been introduced to computationally predict redox-active cysteines<sup>9</sup>, identify cysteines with specific modifications<sup>10–14</sup>, and qualitatively inventory electrophile-modified cysteines in proteomes<sup>15–18</sup>. Some of these studies have provided suggestive evidence that nucleophilic cysteines may possess a variety of important functions<sup>14–18</sup>, although the non-quantitative methods used in each case precluded a robust and systematic evaluation of this potential relationship. We adopted a different strategy to globally characterize cysteine functionality in proteomes based on quantitative reactivity profiling with isotopically labelled, small-molecule electrophiles.

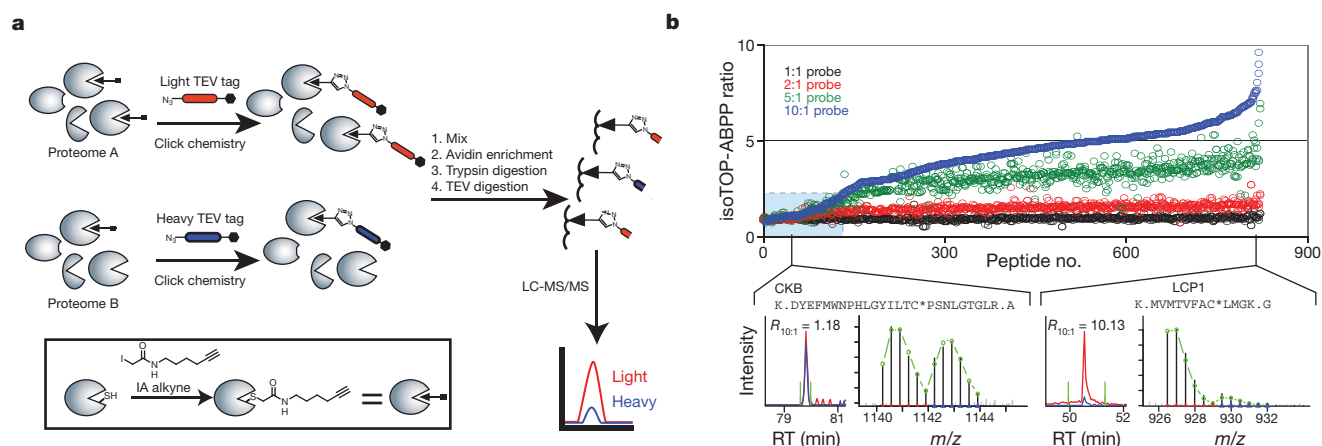
## Quantifying cysteine reactivity in proteomes

Our approach, termed isoTOP-ABPP (isotopic tandem orthogonal proteolysis–activity-based protein profiling), has four features to enable quantitative analysis of native cysteine reactivity (Fig. 1a): (1) an electrophilic iodoacetamide (IA) probe, to label cysteine residues in proteins, that also has (2) an alkyne handle for 'click chemistry' conjugation of probe-labelled proteins<sup>19</sup> to (3) an azide-functionalized TEV-protease recognition peptide containing a biotin group for streptavidin enrichment of probe-labelled proteins<sup>20</sup>, and (4) an isotopically labelled valine for quantitative mass spectrometry (MS) measurements of IA-labelled peptides across multiple proteomes (Supplementary Fig. 1). After tandem on-bead proteolytic digestions with trypsin and TEV protease<sup>15,20</sup>, probe-labelled peptides attached to isotopic tags are released and analysed by liquid-chromatography-high-resolution MS to identify IA-modified cysteines and quantify their extent of labelling based on MS2 and MS1 profiles, respectively. An isoTOP-ABPP ratio,  $R$ , is generated for each identified cysteine that reflects the difference in signal intensity between light and heavy tag-conjugated proteomes.

We first verified the accuracy of isoTOP-ABPP by labelling varying amounts of a mouse liver proteome (1×, 2×, 4×) with the IA probe

<sup>1</sup>The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA. <sup>2</sup>Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA. <sup>3</sup>Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. <sup>4</sup>Interdisciplinary Program in Biomolecular Structure and Design, University of Washington, Seattle, Washington 98195, USA. <sup>5</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.

\*These authors contributed equally to this work.



**Figure 1 | A quantitative approach to globally profile cysteine reactivity in proteomes.** **a**, isoTOP-ABPP involves proteome labelling, click-chemistry-based incorporation of isotopically labelled cleavable tags, and sequential on-bead protease digestions to provide probe-labelled peptides for MS analysis. The IA probe is shown in the inset. LC-MS/MS, liquid-chromatography-MS/MS. **b**, Measured isoTOP-ABPP ratios for peptides from MCF7 cells labelled with four pairwise IA probe concentrations (10:10  $\mu$ M, 20:10  $\mu$ M, 50:10  $\mu$ M, 100:10  $\mu$ M). The blue box highlights peptides with low isoTOP-ABPP ratios ( $R < 2.0$ ). Chromatographs for creatine kinase B (CKB; low ratio) and plastin 2

(LCP1; high ratio) are shown, with elution profiles for heavy- and light-labelled peptides in blue and red, respectively, and green lines depicting peak boundaries used for quantification. Isotopic envelopes are shown for light- and heavy-labelled peptides with green lines representing predicted values. Sequences are shown for tryptic peptides containing IA-probe-labelled cysteines (marked by asterisks) in CKB and LCP1. RT, retention time. Additional chromatographs from isoTOP-ABPP experiments are in Supplementary Table 7.

followed by click chemistry conjugation with either the heavy or light variants of the azide-TEV-biotin tag. The observed signals for labelled cysteines closely matched the expected proteome ratios ( $R_{1:1} \approx 1$ ,  $R_{2:1} \approx 2$ , or  $R_{4:1} \approx 4$ , respectively; Supplementary Fig. 2). A representative MS/MS profile of an IA-labelled peptide from our proteomic experiments is provided in Supplementary Fig. 3.

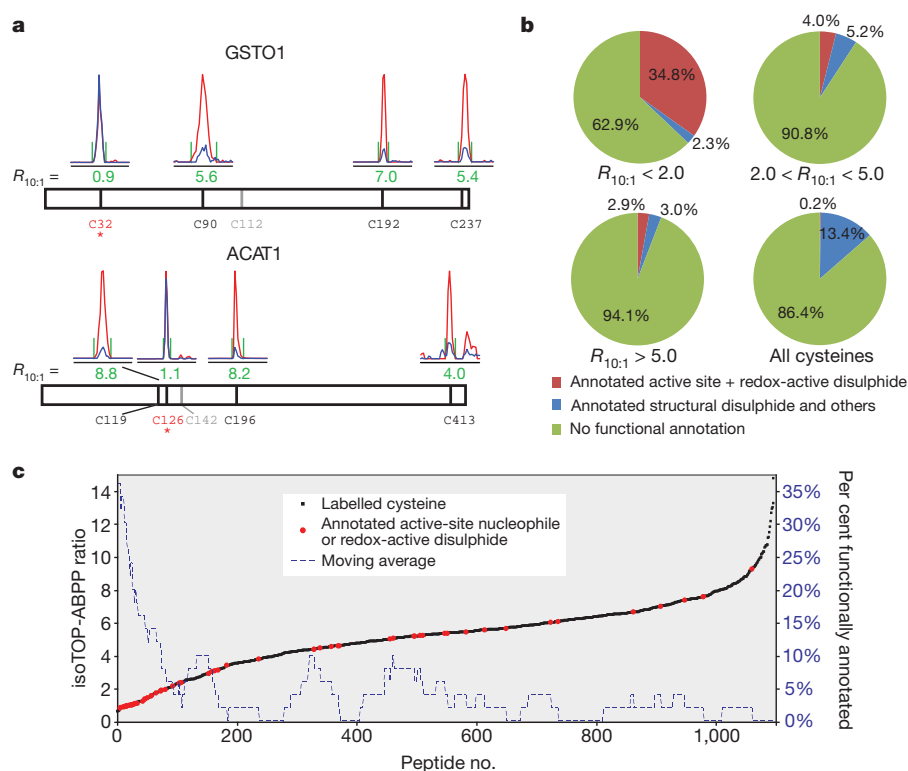
In contrast to traditional cysteine-alkylating protocols for proteomics that use millimolar concentrations of IA to stoichiometrically modify all cysteines in denatured proteins<sup>21</sup>, we proposed that, by applying low (micromolar) concentrations of the IA probe to native proteomes, differences in the extent of alkylation would reflect differences in cysteine reactivity, rather than abundance. This hypothesis predicts that the reactivity of cysteines can be measured on a proteome-wide scale in isoTOP-ABPP experiments that compare low versus high concentrations of IA probe, where hyper-reactive cysteines would be expected to label to completion at low probe concentrations (generating isoTOP-ABPP ratios with  $R_{[\text{high}]:[\text{low}]} \approx 1$ ) and less reactive cysteines should show concentration-dependent increases in IA-probe labelling (generating isoTOP-ABPP ratios with  $R_{[\text{high}]:[\text{low}]} \gg 1$ ) (Supplementary Fig. 4). We tested this idea by performing four parallel isoTOP-ABPP experiments with the soluble proteome of the human breast cancer cell line MCF7 using pair-wise IA-probe concentrations of 10:10  $\mu$ M, 20:10  $\mu$ M, 50:10  $\mu$ M and 100:10  $\mu$ M (light:heavy). More than 800 probe-labelled cysteines were identified on 522 proteins, the vast majority of which exhibited escalating isoTOP-ABPP ratios (Fig. 1b) expected for reactions that did not reach completion over the tested probe concentration range. In contrast, a small subset of cysteines (<10%) showed nearly identical ratios at all probe concentrations tested ( $R_{1:1} \approx R_{2:1} \approx R_{5:1} \approx R_{10:1} \approx 1$ , Fig. 1b, shaded blue box). An expanded analysis of multiple human cancer line (Supplementary Fig. 5 and Supplementary Table 1) and mouse tissue (Supplementary Fig. 6 and Supplementary Table 2) proteomes treated with low (10  $\mu$ M) and high (100  $\mu$ M) IA-probe concentrations revealed consistent isoTOP-ABPP ratios for individual cysteine residues, indicating that the propensity of a cysteine to display high IA reactivity is an intrinsic property of the residue (and presumably its local protein environment), and not, in general, contingent on features specific to a particular cell or tissue. Additionally, isoTOP-ABPP ratios showed no correlation with either protein abundance or peptide ion intensity (Supplementary Fig. 7), indicating that they were

independent of potential MS-based ionization sources for saturation. Finally, we confirmed that similar isoTOP-ABPP ratios were obtained for cysteines in reactions where time rather than the concentration of probe was varied (Supplementary Fig. 8 and Supplementary Table 3), confirming that low isoTOP-ABPP ratios reflect rapid reaction kinetics (hyper-reactivity), rather than saturable binding interactions (see Supplementary Discussion).

### Hyper-reactivity predicts cysteine functionality

We next sought to assess the functional ramifications of the special subset of cysteines that showed hyper-reactivity in isoTOP-ABPP experiments. We first noted that multiple sites of IA-probe labelling on the same protein often showed markedly different isoTOP-ABPP ratios. For example, the glutathione S-transferase GSTO1 was labelled on four cysteine residues, three of which showed high ratios (C90, C192 and C237 had ratios of  $R_{10:1} = 5.6$ , 7, and 5.4, respectively), whereas the fourth (C32) showed a low ratio of  $R_{10:1} = 0.9$  (Fig. 2a). Interestingly, C32 is the active-site nucleophile of GSTO1 (ref. 22). Acetyl-CoA acetyltransferase-1 (ACAT1) was also labelled on four cysteines and three showed high ratios (C119, C196 and C413 showed ratios of  $R_{10:1} = 8.8$ , 8.2 and 4, respectively), whereas the fourth, the active site nucleophile C126 (ref. 23), yielded a low ratio of  $R_{10:1} = 1.1$  (Fig. 2a).

The aforementioned findings indicated that heightened IA reactivity might be a good predictor of cysteine functionality in proteins. To examine this premise more systematically, we queried the Universal Protein Resource (UniProt) database to retrieve functional annotations for the 1,082 cysteine residues labelled by the IA probe. This analysis revealed that the most hyper-reactive cysteines were remarkably enriched in functional residues, with 35% of the cysteines with  $R_{10:1} < 2$  being annotated as active-site nucleophiles or redox-active disulphides compared to 0.2% for all cysteine residues in the UniProt database (Fig. 2b, c, Supplementary Fig. 9 and Supplementary Tables 4 and 5). Hyper-reactive cysteines were also, as a group, more conserved across eukaryotic evolution (Supplementary Fig. 10). A broader survey of hyper-reactive cysteines identified several that have been ascribed functional properties in the literature despite lacking annotation in UniProt (Supplementary Fig. 11). For example, a single hyper-reactive cysteine C108 ( $R_{10:1} = 1.0$ ) was identified in the uncharacterized protein D15Wsu75e. This protein and its orthologues are predicted



**Figure 2 | Hyper-reactive cysteines are highly enriched in functional residues.** **a**, Chromatographs from an isoTOP-ABPP experiment using 100:10  $\mu$ M IA probe are shown for peptides from GSTO1 (top) and ACAT1 (bottom). The cysteine nucleophiles (asterisks) show low ratios ( $R_{10:1} \approx 1$ ), whereas other cysteines show high ratios ( $R_{10:1} \geq 4$ ). **b**, Pie charts illustrating the percentage of functionally annotated cysteines for three isoTOP-ABPP ratio ranges, including an average derived from all cysteines in the UniProt

database. **c**, Correlation of isoTOP-ABPP ratios with functional annotations from the UniProt database where active-site nucleophiles or redox-active disulphides are shown in red, and all other cysteines in black. A moving average (window of 50) of functional residues is shown as a dashed blue line, demonstrating a profound enrichment within  $R_{10:1} < 2.0$ . Data are from experiments in three human cancer cell lines (MCF7, MDA-MB-231 and Jurkat).

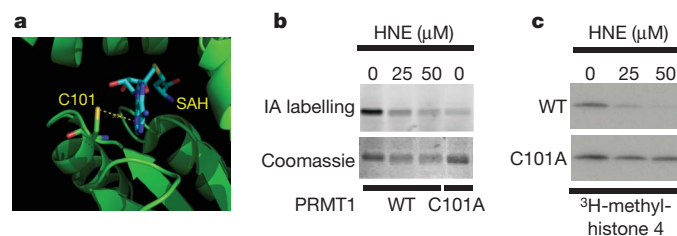
to be cysteine proteases based on conservation of a prototypical Cys-His catalytic dyad<sup>24</sup>. Interestingly, C108 corresponds to the putative cysteine nucleophile of this catalytic motif and a recent crystal structure confirms the proximity of C108 to a conserved histidine (H38) (Supplementary Fig. 12). Thus, quantitative reactivity profiling supports structural predictions that D15Wsu75e is a functional cysteine protease.

Hyper-reactive cysteines also corresponded to sites for post-translational modification. For instance, C101 ( $R_{10:1} = 1.92$ ) in the protein arginine methyltransferase PRMT1 has been identified as a site of modification by the endogenous oxidative product 4-hydroxy-2-nonenal (HNE)<sup>25</sup>. This cysteine, although nonessential for catalytic function, is an active site residue that makes direct contact with the S-adenosylmethionine cofactor<sup>26</sup> (Fig. 3a). Interestingly, we found that HNE inhibited both the IA-labelling (Fig. 3b) and catalytic activity (Fig. 3c) of wild-type PRMT1. A C101A mutant of PRMT1 showed substantially reduced IA-labelling (Fig. 3b) and HNE sensitivity (Fig. 3c). These data indicate that PRMT1 may be regulated by oxidative stress pathways through selective HNE modification of its hyper-reactive, active-site C101 residue. Additional hyper-reactive cysteines represented sites for glutathionylation<sup>27</sup> (CLIC1 (C24), CLIC3 (C25) and CLIC4 (C35);  $R_{10:1} = 2.02$ , 1.07 and 1.45, respectively) and nitrosylation<sup>28</sup> (RTN3; C42,  $R_{10:1} = 0.78$ ). These data, taken together, indicate that heightened reactivity is not only a feature of catalytic cysteines, but also of 'non-catalytic', active-site cysteines, as well as those that undergo various forms of oxidative modification.

### Function of the hyper-reactive cysteine in FAM96B

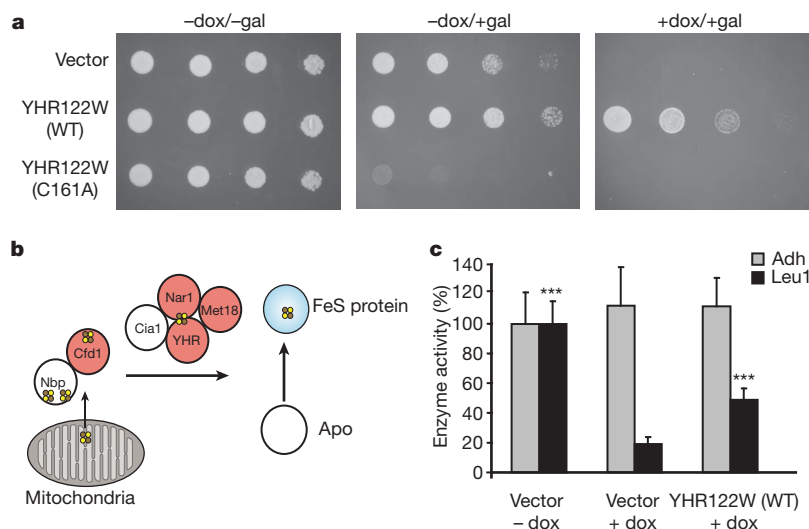
Intrigued by the diverse functional properties showed by hyper-reactive cysteines, we reasoned that critical activities might be inferred for such residues in hitherto uncharacterized proteins. A survey of the cysteines

displaying low isoTOP-ABPP ratios uncovered the highly conserved C93 ( $R_{10:1} = 1.15$ ) in the uncharacterized protein FAM96B (Supplementary Fig. 13). FAM96B has close orthologues in many organisms including the YHR122W protein from the budding yeast *Saccharomyces cerevisiae*, which shows 52% identity with human FAM96B, including conservation of C93 (the corresponding residue in YHR122W is C161). The gene encoding YHR122W is essential for yeast viability<sup>29</sup>, and we found that expression of wild-type YHR122W, but not the C161A mutant of YHR122W could rescue a yeast strain in which the *YHR122W* gene was conditionally suppressed (Fig. 4a and



**Figure 3 | Functional characterization of the hyper-reactive cysteines in PRMT1.** **a**, Crystal structure of rat PRMT1<sup>26</sup> (green, PDB accession code 1ORI) showing the hyper-reactive cysteine C101 in contact with an S-adenosylhomocysteine (SAH) cofactor (cyan). **b**, Wild type (WT) and C101A mutant of human PRMT1 were labelled with the IA probe, followed by click chemistry to incorporate a fluorescent rhodamine tag. In-gel fluorescence demonstrates robust labelling of the wild-type but not C101A mutant PRMT1, and shows that IA-probe labelling of wild-type PRMT1 is inhibited by HNE (upper panel). Lower panel shows Coomassie blue staining for treated protein samples. **c**, Catalytic activity of purified wild-type, but not C101A mutant PRMT1 is inhibited by HNE as measured by monitoring transfer of <sup>3</sup>H-methyl from <sup>3</sup>H-S-adenosylmethionine (SAM) to a histone 4 substrate.





**Figure 4 | Functional characterization of YHR122W/FAM96B.**

**a**, Expression of wild type and a C161A mutant of YHR122W in a yeast strain with a doxycycline (dox)-repressible *YHR122W* gene demonstrated a dominant-negative phenotype on induction of the C161A mutant expression (–dox/+gal, middle panel) and rescue of viability by expression of wild type, but not the C161A mutant of YHR122W (+dox/+gal, right panel). **b**, The cytosolic FeS cluster assembly pathway contains multiple proteins with hyper-reactive cysteines (in red). YHR122W/FAM96B (YHR) is a putative member of

Supplementary Fig. 14). These data confirm the importance of C161 for the *in vivo* function of YHR122W and, by extension, other members of the FAM96B family.

We also observed that expression of the C161A mutant of YHR122W caused a severe growth defect in non-suppressive media indicative of a dominant-negative phenotype (Fig. 4a and Supplementary Fig. 14). This result indicates that the YHR122W protein may engage in protein complexes that are sequestered by the C161A mutant, thereby disrupting the activity of the wild-type protein. Consistent with this premise, queries of the *Saccharomyces* genome databank (SGD) revealed that YHR122W has been found in several large-scale protein interaction studies to bind to proteins involved in cytosolic iron-sulphur (FeS) cluster assembly, namely Nar1 and Cia1 (ref. 30; Fig. 4b). We found that the activity of the FeS-client protein isopropylmalate isomerase (Leu1)<sup>31</sup> was markedly reduced in YHR122W-deleted yeast, and this reduction was substantially rescued by expression of the wild-type YHR122W protein (Fig. 4c). These data support a role for the YHR122W/FAM96B protein in FeS-protein biogenesis. We also note that reactive cysteines seem to be a common feature of proteins in the FeS-protein assembly complex, including the human orthologues of Nar1, Met18 and Cfd1 (NARF, MMS19 and NUBP2, respectively) ( $R_{10:1} = 0.91, 2.2$  and  $2.9$  respectively) (Supplementary Fig. 11), where they may assist in the transfer of assembled FeS clusters to client proteins<sup>32</sup>.

### Predicting functional cysteines in designed proteins

The marked correlation between cysteine hyper-reactivity and functionality observed in native proteomes led us to ask whether this relationship would extend to *de novo* designed proteins. We compared the IA labelling of twelve proteins that were computationally designed to act as cysteine hydrolases. These proteins originated from structurally distinct scaffolds and were all designed to contain cysteine-histidine dyads within an active site cavity (see Supplementary Methods for more details). Two of the designed proteins, ECH13 and ECH19, showed significant hydrolytic activity using a fluorogenic ester substrate, whereas the other ten designs were inactive (Fig. 5a and Supplementary Fig. 15a).

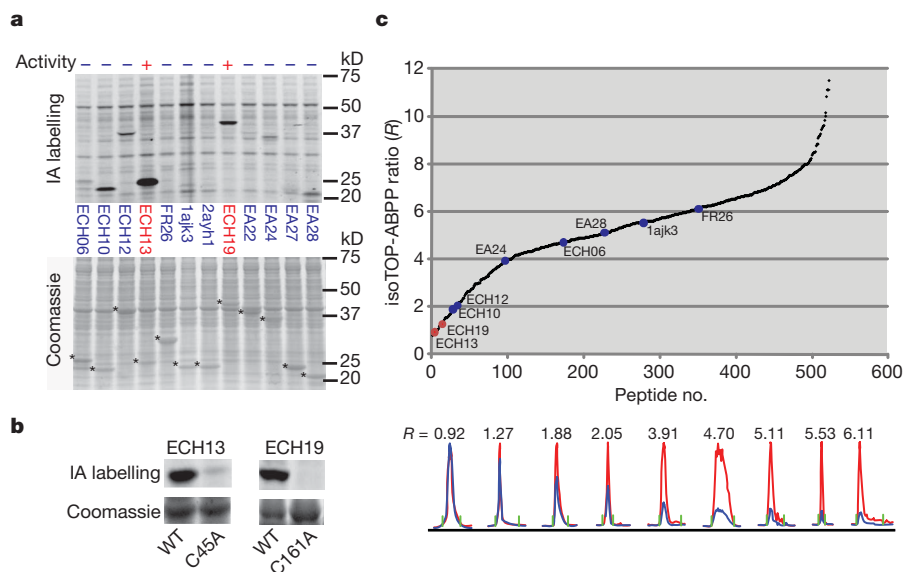
We first evaluated IA labelling of protein designs using a clickable, fluorescent reporter tag and SDS–polyacrylamide gel electrophoresis

this network based on protein–protein interaction studies (see <http://www.yeastgenome.org/>). This panel was adapted from ref. 30. **c**, Doxycycline treatment of the YHR122W-repressible yeast strain significantly decreased the activity of the cytosolic FeS enzyme Leu1<sup>41</sup>, and this activity is rescued by overexpression of wild-type YHR122W. These treatments had no effect on the activity of the non-FeS enzyme alcohol dehydrogenase (ADH). Error bars represent standard deviation,  $n = 3$ . \*\*\* $P < 0.001$ , Student's *t*-test.

(SDS–PAGE) analysis, where similar amounts of each protein were tested in a homogeneous background proteome representing a mix of *Escherichia coli* and human (MCF7 cell line) proteins. The two active protein designs ECH13 and ECH19 showed strong IA-labelling signals compared to inactive designs (Fig. 5a), and, in both cases, mutation of the active-site cysteine to alanine abolished labelling (Fig. 5b) and hydrolytic activity (data not shown). We next combined the proteomes containing all twelve protein designs, diluted them into a background human cell proteome, and analysed the mixture by isoTOP-ABPP. Notably, both ECH13 and ECH19 showed isoTOP-ABPP ratios that were equivalent to the most hyper-reactive cysteines in human and *E. coli* proteomes ( $R_{10:1} = 0.92$  and  $1.27$ , respectively), whereas the remaining inactive protein designs all showed higher ratios ranging from  $1.88$ – $6.11$  (Fig. 5c and Supplementary Fig. 15b, c). These data thus reveal a strong correlation between cysteine hyper-reactivity and hydrolytic activity across a diverse panel of protein designs and designate heightened cysteine nucleophilicity as a key feature of successful cysteine hydrolase designs.

### Conclusions

Here, we have described a quantitative method to profile the intrinsic reactivity of cysteine residues in native proteomes. Measurement of the rate of alkylation by IA (or other carbon electrophiles) has been used by enzymologists to assess the nucleophilicity of cysteine residues in individual, purified proteins<sup>6</sup>. With isoTOP-ABPP, these studies can now be extended to quantitative, proteome-wide surveys of cysteine reactivity in complex biological systems. A key advantage of isoTOP-ABPP over more traditional proteomic methods that target cysteine-containing peptides<sup>14,18</sup> is the use of an alkynylated IA probe in place of more bulky biotinylated reagents, which have shown an impaired ability to label cysteines in native proteins<sup>14</sup>. Alkynylated IA probes, owing to their cell permeability, also afford the opportunity to perform cysteine reactivity profiling in living systems. In pilot experiments, we have found that a large fraction of hyper-reactive cysteines are labelled by the IA probe in living cells (Supplementary Fig. 16). Furthermore, isoTOP-ABPP selectively targets probe-accessible cysteines in native proteins. In this way, structural cysteines engaged in disulphide bonds or buried within the body of a protein are avoided to provide preferential access to a specific fraction of cysteines that are profoundly enriched in functionality (the



**Figure 5 | Quantitative reactivity profiling predicts functional cysteines in designed proteins.** **a**, In-gel fluorescence demonstrates robust IA labelling of two active cysteine hydrolases, ECH13 and ECH19, relative to inactive designs (top panel). Hydrolysis activities of ECH13 and ECH19 measured as the ratio of velocities in the presence versus the absence of purified enzymes were  $71.64 \pm 6.94$  and  $104.15 \pm 10.78$ , respectively (see Supplementary Fig. 15a for substrate hydrolysis assay). Other designs showed no measurable hydrolysis

activity over background ( $0.76 \pm 0.058 \text{ nmol s}^{-1}$ ). Asterisks designate Coomassie blue signals for protein designs (lower panel). **b**, IA labelling is observed for ECH13 and ECH19, but not their active-site cysteine mutants C45A and C161A, respectively. **c**, Catalytic cysteines in ECH13 and ECH19 show low isoTOP-ABPP ratios (red) compared with other designs (blue). Chromatographs are shown for peptides from the nine designs identified in this experiment (bottom panel), in the same order as shown in the top panel.

IA probe labelled 1,082 out of a total of 8,910 cysteines present on the 890 human proteins detected in this study). Projecting forward, it is possible that, by varying the nature of the electrophile, isoTOP-ABPP probes can be created that profile the reactivity of different subsets of cysteines, as well as other amino acids in proteomes, such as serine, threonine, tyrosine and glutamate/aspartate, which have also been shown to react with small-molecule probes<sup>16,18,33–35</sup>.

We discovered that hyper-reactivity can predict cysteine function in both native and designed proteins. The fact that hyper-reactivity was strongly correlated with catalytic activity in *de novo* designed cysteine hydrolases is interesting from the principles of both enzyme engineering and assay development, as it indicates that heightened cysteine nucleophilicity is a key feature of active catalysts and, accordingly, electrophile reactivity could serve as an effective primary screen for novel cysteine-dependent enzymes. We show that these screens can be performed directly in complex proteomes using either gel or MS (isoTOP-ABPP) detection platforms, thus offering a versatile and relatively high-throughput way to evaluate many protein designs in parallel. The isoTOP-ABPP platform has the additional advantage of reading out the relative cysteine reactivity of designs independent of their expression levels against a ‘background’ of native, hyper-reactive cysteines for comparison. isoTOP-ABPP might also offer a complementary way to perform cysteine reactivity/accessibility experiments that monitor protein stability and ligand interactions<sup>36,37</sup>.

The relationship between cysteine reactivity and functionality extends beyond nucleophilic catalysis to include other enzymatic activities (oxidative/reductive), as well as sites of electrophilic and oxidative modification. Quantitative reactivity profiling thus distinguishes itself as a complementary and perhaps more inclusive strategy to survey cysteine function compared to previous computational<sup>9</sup> and experimental<sup>11–14,17</sup> methods that focus on specific cysteine-based activities or modification events. Considering further that hyper-reactive cysteines corresponded to sites for glutathionylation<sup>27</sup>, nitrosylation<sup>28</sup> and HNE-modification<sup>25</sup>, we speculate that cysteine nucleophilicity is a property that may have been selected for during evolution to offer points of protein control by oxidative stress pathways. Determining how the reactivity of cysteine residues is honed will require further investigation, but we anticipate

that quantitative proteomic data, when integrated with the output of ongoing structural genomics programs, may eventually uncover unifying mechanistic principles that explain cysteine reactivity in proteins. In this regard, it is interesting to note that, although hyper-reactive cysteines did not conform to any obvious consensus sequence motifs, many of these residues were found at the N termini of  $\alpha$ -helices (Supplementary Fig. 17). This finding is consistent with literature reports ascribing a role for  $\alpha$ -helix dipoles in the stabilization of cysteine thiolate anions<sup>38</sup>.

Finally, it is important to stress that some functional cysteines may be inherently reactive, but inaccessible to our IA probe for steric reasons. Other cysteine-reactive electrophilic probes<sup>16,17</sup> may prove more suitable for such cysteine residues. Also, hyper-reactivity is not necessarily a defining feature for all functional cysteines. Some enzymes with catalytic cysteines may, for instance, show reduced reactivity until they bind their physiological substrates or may rely more on substrate recognition than inherent catalytic power for function. This may be the case with the E1-activating and E2-conjugating enzymes, which recognize a specific class of ubiquitinated substrates and possess active-site cysteines that showed only moderate levels of electrophile reactivity (Supplementary Fig. 18). Other cysteines may have activities that are not dependent on their nucleophilicity. Our data do indicate, however, that those cysteines that are hyper-reactive in proteomes probably perform important catalytic and/or regulatory functions for their parent proteins. The large number of newly discovered residues that fall into this category foretell a broad role for hyper-reactive cysteines in mammalian biology.

## METHODS SUMMARY

**Probes and tags.** The IA probe and the light and heavy variants of the azide-TEV-biotin tags were synthesized as previously described<sup>20,39</sup>.

**Sample preparation, mass spectrometry and data analysis.** For concentration-dependent experiments, proteome samples in PBS were probe labelled with the desired probe concentration for 1 h. Click chemistry was performed with either the light or heavy variants of the azide-TEV-biotin tags and the samples were mixed and subjected to streptavidin enrichment and subsequent trypsin and TEV digestion. The resulting TEV digests were analysed by Multidimensional Protein Identification Technology (MudPIT) on an LTQ-Orbitrap instrument. The resulting tandem MS data were searched using the SEQUEST algorithm<sup>40</sup> using

a concatenated target/decoy variant of the human, mouse and *E. coli* protein sequence databases. Quantification of light:heavy ratios (isoTOP-ABPP ratios, *R*) was performed using in-house software. Detailed information on sample preparation, mass spectrometry methods and data analysis is presented in Methods.

**Complementation of *S. cerevisiae* YHR122W deletion mutant.** Complementary DNA encoding wild-type YHR122W was subcloned into the pESC\_Leu vector (Stratagene). The YHR122W(C161A) mutant was generated using the Quickchange procedure (Stratagene). These constructs were introduced into a yeast Tet promoter Hughes (yTHC) strain harbouring a conditional (doxycycline-dependent) disruption in the *YHR122W* gene (Open Biosystems). Growth of these transformed cell lines on  $\pm$ gal/ $\pm$ dox media was monitored for 3 days. These cell lines were also used to monitor Leu1 and alcohol dehydrogenase (ADH) activity. Detailed information on the protocols used to subclone, transform and monitor the growth of the yeast strains and measure enzyme activity is available in Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 21 May; accepted 3 September 2010.**

**Published online 21 November 2010.**

- Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. Protein function in the post-genomic era. *Nature* **405**, 823–826 (2000).
- Zhao, Y. & Jensen, O. N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **9**, 4632–4641 (2009).
- Bulaj, G., Kortemme, T. & Goldenberg, D. P. Ionization reactivity relationships for cysteine thiols in polypeptides. *Biochemistry* **37**, 8965–8972 (1998).
- Giles, N. M., Giles, G. I. & Jacob, C. Multiple roles of cysteine in biocatalysis. *Biochem. Biophys. Res. Commun.* **300**, 1–4 (2003).
- Reddie, K. G. & Carroll, K. S. Expanding the functional diversity of proteins through cysteine oxidation. *Curr. Opin. Chem. Biol.* **12**, 746–754 (2008).
- Voss, A. A., Lango, J., Ernst-Russell, M., Morin, D. & Pessah, I. N. Identification of hyperreactive cysteines within ryanodine receptor type 1 by mass spectrometry. *J. Biol. Chem.* **279**, 34514–34520 (2004).
- Lewis, C. T., Seyer, J. M. & Carlson, G. M. Cysteine 288: an essential hyperreactive thiol of cytosolic phosphoenolpyruvate carboxykinase (GTP). *J. Biol. Chem.* **264**, 27–33 (1989).
- Knowles, J. R. Intrinsic  $pK_a$ -values of functional-groups in enzymes: improper deductions from pH-dependence of steady-state parameters. *CRC Crit. Rev. Biochem.* **4**, 165–173 (1976).
- Fomenko, D. E., Xing, W., Adair, B. M., Thomas, D. J. & Gladyshev, V. N. High-throughput identification of catalytic redox-active cysteine residues. *Science* **315**, 387–389 (2007).
- Sethuraman, M. *et al.* Isotope-coded affinity tag (ICAT) approach to redox proteomics: identification and quantitation of oxidant-sensitive cysteine thiols in complex protein mixtures. *J. Proteome Res.* **3**, 1228–1233 (2004).
- Baty, J. W., Hampton, M. B. & Winterbourn, C. C. Proteomic detection of hydrogen peroxide-sensitive thiol proteins in Jurkat cells. *Biochem. J.* **389**, 785–795 (2005).
- Salsbury, F. R., Jr, Knutson, S. T., Poole, L. B. & Fetrow, J. S. Functional site profiling and electrostatic analysis of cysteines modifiable to cysteine sulfenic acid. *Protein Sci.* **17**, 299–312 (2008).
- Leonard, S. E., Reddie, K. G. & Carroll, K. S. Mining the thiol proteome for sulfenic acid modifications reveals new targets for oxidation in cells. *ACS Chem. Biol.* **4**, 783–799 (2009).
- Kim, J.-R., Yoon, H. W., Kwon, K.-S., Lee, S.-R. & Rhee, S. G. Identification of proteins containing cysteine residues that are sensitive to oxidation by hydrogen peroxide at neutral pH. *Anal. Biochem.* **283**, 214–221 (2000).
- Speers, A. E. & Cravatt, B. F. A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *J. Am. Chem. Soc.* **127**, 10018–10019 (2005).
- Weerapana, E., Simon, G. M. & Cravatt, B. F. Disparate proteome reactivity profiles of carbon electrophiles. *Nature Chem. Biol.* **4**, 405–407 (2008).
- Dennehy, M. K., Richards, K. A., Wernke, G. R., Shyr, Y. & Liebler, D. C. Cytosolic and nuclear protein targets of thiol-reactive electrophiles. *Chem. Res. Toxicol.* **19**, 20–29 (2006).
- Shin, N.-Y., Liu, Q., Stamer, S. L. & Liebler, D. C. Protein targets of reactive electrophiles in human liver microsomes. *Chem. Res. Toxicol.* **20**, 859–867 (2007).
- Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling *in vivo* using a copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686–4687 (2003).
- Weerapana, E., Speers, A. E. & Cravatt, B. F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)—a general method for mapping sites of probe modification in proteomes. *Nature Protocols* **2**, 1414–1425 (2007).
- Shiio, Y. & Aebersold, R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nature Protocols* **1**, 139–145 (2006).
- Board, P. G. *et al.* Identification, characterization, and crystal structure of the omega class glutathione transferases. *J. Biol. Chem.* **275**, 24798–24806 (2000).
- Thompson, S. *et al.* Mechanistic studies on  $\beta$ -ketoacyl thiolase from *Zoogloea ramigera*: identification of the active-site nucleophile as Cys89, its mutation to Ser89, and kinetic and thermodynamic characterization of wild-type and mutant enzymes. *Biochemistry* **28**, 5735–5742 (1989).
- Iyer, L. M., Koonin, E. V. & Aravind, L. Novel predicted peptidases with a potential role in the ubiquitin signaling pathway. *Cell Cycle* **3**, 1440–1450 (2004).
- Codreanu, S. G., Zhang, B., Sobocki, S. M., Billheimer, D. D. & Liebler, D. C. Global analysis of protein damage by the lipid electrophile 4-hydroxy-2-nonenal. *Mol. Cell. Proteomics* **8**, 670–680 (2009).
- Zhang, X. & Cheng, X. Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. *Structure* **11**, 509–520 (2003).
- Harrop, S. J. *et al.* Crystal structure of a soluble form of the intracellular chloride ion channel CLIC2 at 1.4-Å resolution. *J. Biol. Chem.* **276**, 44993–45000 (2001).
- Hao, G., Derakhshan, B., Shi, L., Campagne, F. & Gross, S. S. SNOSID, a proteomic method for identification of cysteine S-nitrosylation sites in complex protein mixtures. *Proc. Natl Acad. Sci. USA* **103**, 1012–1017 (2006).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Lill, R. Function and biogenesis of iron sulphur proteins. *Nature* **460**, 831–838 (2009).
- Pierik, A. J., Netz, D. J. & Lill, R. Analysis of iron–sulfur protein maturation in eukaryotes. *Nature Protocols* **4**, 753–766 (2009).
- Netz, D. J. A., Pierik, A. J., Stumpf, M., Muhlenhoff, U. & Lill, R. The Cfd1–Nbp35 complex acts as a scaffold for iron–sulfur protein assembly in the yeast cytosol. *Nature Chem. Biol.* **3**, 278–286 (2007).
- Okerberg, E. S. *et al.* High-resolution functional proteomics by active-site peptide profiling. *Proc. Natl Acad. Sci. USA* **102**, 4996–5001 (2005).
- Nazif, T. & Bogoy, M. Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proc. Natl Acad. Sci. USA* **98**, 2967–2972 (2001).
- Chen, G. *et al.* Reactivity of functional groups on the protein surface: development of epoxide probes for protein labeling. *J. Am. Chem. Soc.* **125**, 8130–8133 (2003).
- Silverman, J. A. & Harbury, P. B. Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J. Biol. Chem.* **277**, 30968–30975 (2002).
- Isom, D. G., Vardy, E., Oas, T. G. & Helling, H. W. Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity. *Proc. Natl Acad. Sci. USA* **107**, 4908–4913 (2010).
- Kortemme, T. & Creighton, T. E. Ionisation of cysteine residues at the termini of model  $\alpha$ -helical peptides. Relevance to unusual thiol  $pK_a$  values in proteins of the thioredoxin family. *J. Mol. Biol.* **253**, 799–812 (1995).
- Macpherson, L. J. *et al.* Noxious compounds activate TRPA1 ion channels through covalent modification of cysteines. *Nature* **445**, 541–545 (2007).
- Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
- Kispal, G., Csere, P., Prohl, C. & Lill, R. The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J.* **18**, 3981–3989 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank T. Bartfai, I. Wilson and members of the B.F.C. laboratory for comments and critical reading of the manuscript, T. Ji for experimental assistance and J. Gallaher for expression of designed proteins. This work was supported by the National Institutes of Health (CA087660, MH084512), a Pfizer Postdoctoral Fellowship (E.W.), a Koshland Graduate Fellowship in Enzyme Biochemistry (G.M.S.), a National Science Foundation predoctoral fellowship (D.A.B.) and the Skaggs Institute for Chemical Biology.

**Author Contributions** B.F.C., E.W. and C.W. conceived the project and E.W. and C.W. performed MS experiments and yeast growth/Leu1 assays. C.W. and G.M.S. performed computational data analyses. S.K., F.R. and D.B. performed computational design of cysteine hydrolases and measured activity using a fluorogenic assay. D.A.B. purified PRMT1 and M.B.D.D. and K.M. performed PRMT1 activity assays. B.F.C., E.W., C.W. and G.M.S. analysed data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to B.F.C. ([cravatt@scripps.edu](mailto:cravatt@scripps.edu)).



## METHODS

All compounds and reagents were purchased from Novabiochem, Sigma or Fisher, except where noted.

**Preparation of mouse proteomes.** Mouse tissues (heart and liver) were harvested and immediately flash frozen in liquid nitrogen. The tissues were then Dounce homogenized in 1× PBS, pH 7.4. Centrifugation at 100,000g (45 min) provided soluble fractions (supernatant) and membrane fractions (pellet). Protein concentrations for each proteome were obtained using the Bio-Rad DC protein assay and stored at  $-80^{\circ}\text{C}$  till use.

**Preparation of human cancer cell line proteomes.** MDA-MB-231 cells were grown in L15 media supplemented with 10% fetal bovine serum at  $37^{\circ}\text{C}$  in a  $\text{CO}_2$ -free incubator. Jurkat cells and MCF7 cells were grown in RPMI-1640 supplemented with 10% fetal bovine serum at  $37^{\circ}\text{C}$  with 5%  $\text{CO}_2$ . For *in vitro* labelling experiments, cells were grown to 100% confluency, washed three times with PBS and scraped in cold PBS. Cell pellets were isolated by centrifugation at 1,400g for 3 min, and the cell pellets stored at  $-80^{\circ}\text{C}$  until further use. For *in situ* labelling of MDA-MB-231 and MCF7 cells, the cells were grown to 90% confluency, the media was removed and replaced with fresh media containing 10  $\mu\text{M}$  IA probe. The cells were incubated at  $37^{\circ}\text{C}$  for 1 h and harvested as detailed above. The harvested cell pellets were lysed by sonication and fractionated by centrifugation (100,000g, 45 min) to yield soluble and membrane proteomes. The proteomes were diluted to 2  $\text{mg ml}^{-1}$  and stored at  $-80^{\circ}\text{C}$  until use.

**Protein labelling and click chemistry.** Proteome samples were diluted to a 2  $\text{mg protein/ml}$  solution in PBS. Each sample (2 × 0.5 ml aliquots) was treated with 10, 20, 50, or 100  $\mu\text{M}$  of IA probe using 5  $\mu\text{l}$  of a 1, 2, 5, or 10 mM stock in DMSO. The labelling reactions were incubated at room temperature ( $25^{\circ}\text{C}$ ) for 1 h. Click chemistry was performed by the addition of 150  $\mu\text{M}$  of either the light TEV tag or heavy TEV tag (15  $\mu\text{l}$  of a 5 mM stock), 1 mM tris(2-carboxyethyl)phosphine (TCEP; fresh 50× stock in water), 100  $\mu\text{M}$  ligand (17× stock in DMSO: *t*-butanol 1:4) and 1 mM  $\text{CuSO}_4$  (50× stock in water). Samples were allowed to react at room temperature for 1 h. After the click chemistry step, the light- and heavy-labelled samples were mixed together and centrifuged (5,900g, 4 min,  $4^{\circ}\text{C}$ ) to pellet the precipitated proteins. The pellets were washed twice in cold MeOH, after which the pellet was solubilized in PBS containing 1.2% SDS via sonication and heating (5 min,  $80^{\circ}\text{C}$ ).

For time course experiments, proteome samples were labelled with 100  $\mu\text{M}$  of IA probe (using 5  $\mu\text{l}$  of a 10 mM stock in DMSO). After 6 min of probe labelling, an aliquot of the reaction was quenched by passing the sample through a NAP-5 column (GE Healthcare) to remove excess, unreacted probe. After 60 min of probe labelling, the other sample was quenched as before and click chemistry was performed as described earlier.

**Streptavidin enrichment of probe-labelled proteins.** The SDS-solubilized, probe-labelled proteome samples were diluted with 5 ml of PBS for a final SDS concentration of 0.2%. The solutions were then incubated with 100  $\mu\text{l}$  of streptavidin-agarose beads (Pierce) for 3 h at room temperature. The beads were washed with 10 ml 0.2% SDS/PBS, 3 × 10 ml PBS and 3 × 10 ml  $\text{H}_2\text{O}$  and the beads were pelleted by centrifugation (1,300g, 2 min) between washes.

**On-bead trypsin and TEV digestion.** The washed beads described earlier were suspended in 500  $\mu\text{l}$  of 6 M urea/PBS and 10 mM TCEP (from 20× stock in  $\text{H}_2\text{O}$ ) and placed in a  $65^{\circ}\text{C}$  heat block for 15 min. Twenty millimolar iodoacetamide (from 50× stock in  $\text{H}_2\text{O}$ ) was then added and allowed to react at  $37^{\circ}\text{C}$  for 30 min. Following reduction and alkylation, the beads were pelleted by centrifugation (1,300g, 2 min) and resuspended in 200  $\mu\text{l}$  of 2 M urea/PBS, 1 mM  $\text{CaCl}_2$  (100× stock in  $\text{H}_2\text{O}$ ), and trypsin (2  $\mu\text{g}$ ). The digestion was allowed to proceed overnight at  $37^{\circ}\text{C}$ . The digest was separated from the beads using a Micro Bio-Spin column and the beads were then washed with 3 × 500  $\mu\text{l}$  PBS, 3 × 500  $\mu\text{l}$   $\text{H}_2\text{O}$ , and 1 × 150  $\mu\text{l}$  of TEV digest buffer. The washed beads were then resuspended in 150  $\mu\text{l}$  of TEV digest buffer with AcTEV Protease (Invitrogen, 5  $\mu\text{l}$ ) for 12 h at  $29^{\circ}\text{C}$ . The eluted peptides were separated from the beads using a Micro Bio-Spin column and the beads washed with  $\text{H}_2\text{O}$  (2 × 75  $\mu\text{l}$ ). Formic acid (15  $\mu\text{l}$ ) was added to the sample, which was stored at  $-20^{\circ}\text{C}$  until MS analysis.

**Liquid-chromatography-mass-spectrometry (LC-MS) analysis.** LC-MS/MS analysis was performed on an LTQ-Orbitrap mass spectrometer (ThermoFisher) coupled to an Agilent 1100 series high-performance liquid chromatography system. TEV digests were pressure loaded onto a 250  $\mu\text{m}$  fused silica desalting column packed with 4 cm of Aqua C18 reverse phase resin (Phenomenex). The peptides were then eluted onto a biphasic column (100  $\mu\text{m}$  fused silica with a 5  $\mu\text{m}$  tip, packed with 10 cm C18 and 3 cm Partisphere strong cation exchange resin (SCX, Whatman) using a gradient 5–100% buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B: 20% water, 80% acetonitrile, 0.1% formic acid). The peptides were then eluted from the SCX onto the C18 resin and into the mass spectrometer using four salt steps as previously described<sup>15,20</sup>. The flow rate

through the column was set to  $\sim 0.25 \mu\text{l min}^{-1}$  and the spray voltage was set to 2.75 kV. One full MS scan (FTMS) (400–1,800 MW) was followed by 18 data dependent scans (ITMS) of the *n*th most intense ions with dynamic exclusion disabled.

**Peptide identification.** The tandem MS data were searched using the SEQUEST algorithm<sup>40</sup> using a concatenated target/decoy variant of the human and mouse International Protein Index databases. A static modification of +57.02146 on cysteine was specified to account for iodoacetamide alkylation and differential modifications of +464.28596 (light probe modification) and +470.29977 (heavy probe modification) were specified on cysteine to account for probe modifications with the either light or heavy variants of the IA-probe-TEV adduct. SEQUEST output files were filtered using DTASelect 2.0<sup>42</sup>. Reported peptides were required to be fully tryptic and contain the desired probe modification and discriminant analyses were performed to achieve a peptide false-positive rate below 5%. The actual false-positive rate was assessed at this stage according to established guidelines<sup>43</sup> and found to be  $\sim 3.5\%$ . Additional assessments of the false-positive rate were performed following the application of additional filters (described later) resulting in a final false-positive rate below 0.05%.

**Ratio quantification.** Quantification of light/heavy ratios (isoTOP-ABPP ratios, *R*) was performed using in-house software written in the R programming language that utilizes routines from the open-source XCMS package<sup>44</sup> for MS data analysis to read in raw chromatographic data in the mzXML format<sup>45</sup>. Each experiment consisted of two LC/LC-MS/MS runs: light:heavy 10  $\mu\text{M}$ :10  $\mu\text{M}$ , and light:heavy 100  $\mu\text{M}$ :10  $\mu\text{M}$  IA-probe concentration. Both runs were searched using SEQUEST and filtered with DTASelect as described earlier. Because the mass spectrometer was configured for data-dependant fragmentation, peptides are not always identified in every run. As such, peptides were identified in either 1) only the 10  $\mu\text{M}$ :10  $\mu\text{M}$  run, 2) only the 100  $\mu\text{M}$ :10  $\mu\text{M}$  run, or 3) both runs. In the case of peptides that were sequenced in both runs, identification of the corresponding peaks was made by choosing peaks that co-elute with the peptide identification. In the case of probe-modified peptides that were sequenced in one, but not the other run, an algorithm was developed to identify the corresponding peak in the run without the SEQUEST identification. To accomplish this, the retention time of the 'reference' peptide is used to position a retention time window ( $\pm 10$  min) across the run lacking a peptide identification. Extracted ion chromatograms ( $\pm 10$  p.p.m.) of the target peptide *m/z* with both 'light' and 'heavy' modifications are generated within that window. The program then searches for candidate co-eluting pairs of light:heavy MS1 peaks, and for each candidate pair calculates the ratio of integrated peak area between the light and heavy peaks. Several filters are used to ensure that the correct peak pair is identified. First, the extent of co-elution for each peak pair is quantified using a Pearson correlation, an established method to gauge elution profile similarity<sup>46</sup>. Second, the predicted pattern of the isotopic envelope of the target peptide is generated and compared to the observed high-resolution MS1 spectrum. This comparison generates an 'envelope correlation score' (Env) that also enables confirmation of the monoisotopic mass and charge state of each candidate peak. Peak pairs that have poor co-elution scores, or that have the incorrect monoisotopic mass or charge, or whose isotopic envelopes are not well correlated with the predicted envelope are eliminated from consideration. After application of these filters, in the rare case that multiple candidates still exist, then no peak is chosen and a ratio is not recorded. Usually, however, application of these filters results in a single candidate peak pair and the ratio for this peak pair is recorded for the peptide in the corresponding run. In this way, each experiment yields two ratios, one for the 10  $\mu\text{M}$ :10  $\mu\text{M}$  run and one for the 100  $\mu\text{M}$ :10  $\mu\text{M}$  run. Following application of these filters, the false-positive rate was reassessed, and found to be less than 0.05% in all cases.

After ratios for unique peptide entries are calculated for each experiment, overlapping peptides with the same labelled cysteine (for example, same local sequence around the labelled cysteines but different charge states, MudPIT segment numbers, or tryptic termini) are grouped together, and the median ratio from each group is reported as the final ratio (*R*). All of these values can be found in Supplementary Tables 1, 2 and 3 and representative chromatographs can be seen in Supplementary Table 7. Raw result files of peptide identification using SEQUEST can be found in Supplementary Table 9.

**Functional annotation of labelled cysteines.** For automated functional analyses, custom perl-scripts were developed to query the UniProtKB/Swiss-Prot Protein Knowledgebase release 57.4 (current as of 16 June 2009). Sequence annotation in the (Features) section of the relevant UniProt entry was mined and any annotation corresponding to the labelled residue was collected. This functional annotation in its entirety can be found in Supplementary Tables 4 and 5.

**Recombinant PRMT1 protein expression and purification.** Full-length cDNA encoding human PRMT1 in pOTB7 was purchased from Open BioSystems and subcloned into pET-45b(+) (Novagen). BL21(DE3) *E. coli* containing this vector was grown in LB media containing 75  $\text{mg l}^{-1}$  carbenicillin with shaking at  $37^{\circ}\text{C}$

to an OD<sub>600 nm</sub> of 0.5. The cells were then induced with 1 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG) and harvested 4 h later by centrifugation. Cells were lysed by stirring for 20 min at 4 °C in 50 mM Tris-HCl (pH 8.0) with 150 mM NaCl and supplemented with 1 mg ml<sup>-1</sup> lysozyme and 1 mg ml<sup>-1</sup> DNase I. The lysate was then sonicated and centrifuged at 10,000g for 10 min. Talon cobalt affinity resin (Clontech; 400  $\mu$ l of slurry per gram of cell paste) was added to the supernatant, and the mixture was rotated at 25 °C for 30 min. Beads were collected by centrifugation at 700g for 3 min, washed twice with Tris buffer, and applied to a 1-cm column. The column was washed twice with Tris buffer (10 ml per 400  $\mu$ l of resin slurry) and Tris buffer with 500 mM NaCl once. The bound protein was eluted by the addition of 100 mM imidazole (2 ml per 400  $\mu$ l of resin). Imidazole was removed by passage over a Sephadex G-25M column (GE Healthcare), and the eluate was concentrated using an Amicon centrifugal filter device (Millipore). Protein concentration was determined using the Bio-Rad DC protein assay kit. These conditions yielded PRMT1 at approximately 0.5 mg l<sup>-1</sup> of culture. A C101A mutation was introduced into the pET-45b(+) construct described earlier using the Quickchange Site-Directed Mutagenesis Kit (Stratagene), and the resulting mutant protein was expressed identically and isolated with a similar yield.

**In-gel fluorescence characterization of PRMT1.** Thirteen micrograms of recombinant PRMT1 (wild type or C101A mutant) in 50  $\mu$ l PBS buffer was pre-incubated with 0, 25 or 50  $\mu$ M HNE (Calbiochem, 50 mM stock in ethanol) for 1 h at room temperature and was then labelled with 100 nM of the IA probe (5  $\mu$ M stock in DMSO) and the reactions incubated for 1 h at room temperature. Click chemistry was performed with 20  $\mu$ M rhodamine-azide, 1 mM TCEP, 100  $\mu$ M TBTA ligand and 1 mM CuSO<sub>4</sub>. The reaction was allowed to proceed at room temperature for 1 h before quenching with 50  $\mu$ l of 2 $\times$  SDS-PAGE loading buffer (reducing). Quenched reactions were separated by SDS-PAGE (30  $\mu$ l of sample/lane) and visualized in-gel using a Hitachi FMBio Iie flatbed laser-induced fluorescence scanner (MiraiBio).

**PRMT1 in vitro methylation assays.** Five-hundred nanograms of recombinant human PRMT1 (wild type or C101A mutant) was pre-incubated with HNE (Calbiochem) for 30 min and methylation activity was monitored after addition of 1 mg of recombinant histone 4 (M2504S; NEB) and SAM (2  $\mu$ Ci) in methylation buffer (20 mM Tris, pH 8.0, 200 mM NaCl, 0.4 mM EDTA). Reactions were incubated for 90 min at 30 °C and stopped with SDS sample buffer. SDS-PAGE gels were fixed with 10% acetic acid/10% methanol v/v, washed, and incubated with Amplify reagent (Amersham) before exposing at -80 °C.

**Complementation of *S. cerevisiae* YHR122W deletion mutant.** A cDNA encoding YHR122W was purchased as a full-length expressed sequence tag (Open Biosystems). The construct for subcloning into the yeast epitope tagging vector pESC-Leu (Stratagene) was generated by polymerase chain reaction (PCR) from the corresponding cDNA using the following primers: sense primer, 5'-GAAGCGGCCGCAATGTCTGAGTTTTTGAATGA-3'; antisense primer, 5'-CCGACTAGTGCCTTACAAGTCACTAACATCTTAG-3'.

The PCR product was digested with NotI-SpeI and subcloned into a NotI-SpeI-digested pESC-Leu vector and sequenced. The YHR122W(C161A) mutant was generated using the Quickchange procedure (Stratagene). The mutant cDNA was sequenced and found to contain only the desired mutation.

Constructs containing wild-type and C161A mutant YHR122W were introduced into the yTHC strain YSC1180-7428770 (Open Biosystems) using the reagents provided in the Yeastmaker Yeast Transformation System 2 (Clontech). The yeast was grown in synthetic dextrose minimal medium (-Leu) and spot assays were performed in either synthetic dextrose minimal medium (-Leu) or synthetic galactose minimal medium (-Leu) + agar plates  $\pm$  50  $\mu$ g ml<sup>-1</sup> doxycycline. The plates were cultured at 30 °C for 3 days.

**Isopropylmalate isomerase (Leu1) assay.** Yeast strains harbouring either an empty vector or wild-type YHR122W (see earlier section) were cultured in synthetic dextrose minimal medium (-Leu) to an OD<sub>600 nm</sub> of 1.0 and transferred into synthetic galactose minimal medium (-Leu)  $\pm$  50  $\mu$ g ml<sup>-1</sup> doxycycline for 12 h. Yeast were lysed and Leu1 semi-purified by ammonium sulphate precipitation

(40–70%). The activity assays were performed using DL-threo-3-isopropylmalic acid as the substrate and product formation was measured by monitoring absorbance at 235 nm for 10 min<sup>31</sup>.

**ADH assay.** Yeast cell lysates in 0.1 M sodium pyrophosphate buffer (pH 9.2, 1.5 ml) were treated with 2 M ethanol (0.5 ml) and 0.025 M NAD (1.0 ml) and ADH activity was measured by absorbance increase at 340 nm for 3 min<sup>47</sup>.

**De novo designs of cysteine hydrolases and hydrolysis activity assays.** We used the Rosetta computational enzyme design methodology<sup>48</sup> to search a set of protein scaffolds for constellations of backbones capable of supporting an idealized transition state for ester hydrolysis derived from the geometries and mechanisms of natural cysteine hydrolases<sup>49</sup>. The idealized active-site models feature a nucleophilic cysteine, a general base/acid histidine and at least one side-chain or backbone hydrogen bond donor as the oxyanion hole. The sequence of residues surrounding the putative active sites was optimized using the Rosetta design algorithm to maximize transition state stabilization<sup>50</sup>. A set of 12 designed proteins in 10 distinct scaffolds was chosen for experimental characterization. For each designed protein, synthetic genes were obtained and protein expression and purification was performed in *E. coli* as previously described<sup>50</sup>. Activity was measured with the substrate by following the initial (<5% substrate conversion) increase in fluorescence due to the appearance of the product coumarin. A protein concentration of 20  $\mu$ M and substrate concentration of 100  $\mu$ M were used in 25 mM HEPES buffer, 150 mM NaCl, 1 mM TCEP, pH 7.5. The background rate was measured under identical conditions but without the protein. Kunkel mutagenesis was used for creating point mutations in the active-site residues. A detailed description of the design and characterization of the cysteine hydrolases will be presented elsewhere. Amino acid sequences of the 12 designs can be found in Supplementary Information.

**In-gel fluorescence and isoTOP-ABPP characterization of designed proteins.** For in-gel fluorescence studies, *E. coli* lysates overexpressing the designed proteins were diluted to 2 mg protein/ml in PBS. Each sample (25  $\mu$ l) was mixed with 25  $\mu$ l of MCF7 human cell soluble proteome (2 mg ml<sup>-1</sup>) and was labelled with 100 nM of the IA probe (5  $\mu$ M stock in DMSO) and the reactions incubated for 1 h at room temperature. Click chemistry, SDS-PAGE separation and in-gel fluorescence visualization were performed as described in previous sections.

For isoTOP-ABPP studies, 10  $\mu$ l of each of the *E. coli* lysates (2 mg protein/ml) overexpressing the designed constructs were mixed together and the total volume was brought to 1 ml by the addition of 2 mg ml<sup>-1</sup> of MCF7 soluble proteome. Time-dependent and concentration-dependent labelling with the IA probe, click chemistry, on-bead trypsin and TEV digestions, LC-MS runs and MS data analysis were performed as described in previous sections.

42. Tabb, D. L., McDonald, W. H. & Yates, J. R. III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21–26 (2002).
43. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4**, 207–214 (2007).
44. Collins, S. R. et al. Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **6**, 439–450 (2007).
45. Pedrioli, P. G. A. et al. A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnol.* **22**, 1459–1466 (2004).
46. Park, S. K., Venable, J. D., Xu, T. & Yates, J. R. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nature Methods* **5**, 319–322 (2008).
47. Vallee, B. L. & Hoch, F. L. Zinc, a component of yeast alcohol dehydrogenase. *Proc. Natl Acad. Sci. USA* **41**, 327–338 (1955).
48. Zanghellini, A. et al. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006).
49. Ma, S., Devi-Kesavan, L. S. & Gao, J. Molecular dynamics simulations of the catalytic pathway of a cysteine protease: a combined QM/MM study of human cathepsin K. *J. Am. Chem. Soc.* **129**, 13633–13645 (2007).
50. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).

# Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA

Nicholas J. Reiter<sup>1</sup>, Amy Osterman<sup>1</sup>, Alfredo Torres-Larios<sup>1†</sup>, Kerren K. Swinger<sup>1†</sup>, Tao Pan<sup>2</sup> & Alfonso Mondragón<sup>1</sup>

**Ribonuclease (RNase) P is the universal ribozyme responsible for 5'-end tRNA processing. We report the crystal structure of the *Thermotoga maritima* RNase P holoenzyme in complex with tRNA<sup>Phe</sup>. The 154 kDa complex consists of a large catalytic RNA (P RNA), a small protein cofactor and a mature tRNA. The structure shows that RNA–RNA recognition occurs through shape complementarity, specific intermolecular contacts and base-pairing interactions. Soaks with a pre-tRNA 5' leader sequence with and without metal help to identify the 5' substrate path and potential catalytic metal ions. The protein binds on top of a universally conserved structural module in P RNA and interacts with the leader, but not with the mature tRNA. The active site is composed of phosphate backbone moieties, a universally conserved uridine nucleobase, and at least two catalytically important metal ions. The active site structure and conserved RNase P–tRNA contacts suggest a universal mechanism of catalysis by RNase P.**

Ribonuclease P (RNase P) is a ribonucleoprotein complex responsible for processing many different RNA molecules in the cell (for recent reviews, see refs 1–3). It is found in almost all organisms and is composed of one essential RNA subunit and one or more protein subunits. The RNA component is responsible for catalysis and can process RNA *in vitro* in the absence of protein, albeit with reduced efficiency<sup>4</sup>. The discovery that the RNA component is the catalytic moiety<sup>4</sup> helped cement the notion that RNA can be directly involved in catalysis. RNase P is considered a remnant of an ancient RNA-based world and an example of an RNA-based catalyst with many features in common with protein-based catalysts.

RNase P recognizes its substrate in *trans* and is a multiple turnover enzyme. The preferred substrate is pre-tRNA and recognition involves features distant from the cleavage site, such as the TΨC loop of the tRNA acceptor stem<sup>5</sup>. RNA cleavage requires divalent metals<sup>4,6,7</sup>, yet the chemical mechanism and the location of the active site remain largely undefined as well as the exact role of the protein components. In the case of bacterial RNase P, the single essential protein improves the reaction rate by two to three orders of magnitude<sup>8,9</sup>, helps to stabilize the active P RNA fold<sup>8,10</sup>, binds the 5' leader region of the pre-tRNA substrate<sup>11,12</sup>, and assists in product release<sup>13</sup>.

Structural studies of the RNA component reveal a two domain (S- and C-domains) molecule formed by single and coaxial stems linked together by a variety of tertiary interactions<sup>14–17</sup>, including five conserved regions I to V (CR-I to CR-V) of P RNA that are common to all organisms<sup>18</sup>. These conserved regions cluster into two areas, one involved in substrate recognition and the other forming the active site scaffold<sup>19</sup>.

Here we present the crystal structure of *Thermotoga maritima* RNase P holoenzyme in complex with mature tRNA<sup>Phe</sup>, and also the structure of the complex in the presence of a post-cleavage tRNA leader. The two structures help answer key questions about the mechanism of this crucial ribozyme with implications for a broader understanding of the general mechanisms of RNA–RNA based recognition and catalysis.

## Structure determination

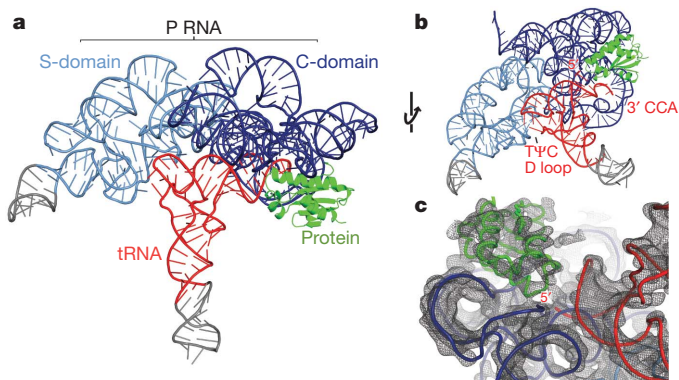
The components of the complex were purified separately and assembled by mixing and heating before crystallization (see Methods). The pre-tRNA was processed into mature tRNA and hence the structure represents a ribozyme–product complex. To promote crystal formation, two interaction modules<sup>20</sup> were introduced, which had a modest effect on catalytic activity (Supplementary Fig. 1 and Supplementary Table 1). The crystals diffract anisotropically to 3.8 Å and ~4.0 Å. An initial 6 Å map was obtained from phases from a Ta<sub>6</sub>Br<sub>12</sub> derivative; these phases helped locate heavy atoms in other derivative data sets. Multiple isomorphous replacement with anomalous scattering (MIRAS) phases produced an excellent map to 4.1 Å where all three components were visible (Fig. 1, Supplementary Figs 2 and 3, and Supplementary Tables 2–4). Density was particularly clear for the RNA molecules, whereas density was only clear for the protein backbone and hence the high resolution model of the *T. maritima* protein<sup>21</sup> was positioned without significant rebuilding. The P RNA was built into the map using the structures of *T. maritima*<sup>17</sup> and *Bacillus stearothermophilus*<sup>14</sup> P RNA as guides, whereas *T. maritima* tRNA<sup>Phe</sup> used yeast tRNA<sup>Phe</sup> as a guide<sup>22</sup>. The structure was refined using anisotropic data to 3.8 Å resolution. Crystals with a tRNA leader present were obtained by soaking a short oligonucleotide with and without samarium chloride and this structure was refined to 4.2 Å resolution.

## Overall structure

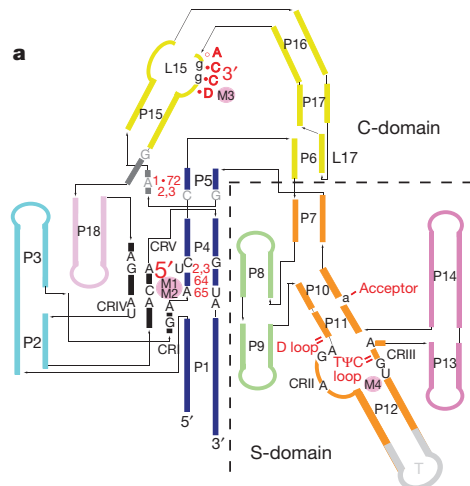
In the complex, the tRNA sits with the acceptor stem against RNase P, making several tRNA–P RNA intermolecular contacts (Fig. 1 and Supplementary Fig. 1). The TΨC and D loops of the tRNA contact the S-domain, while the acceptor stem extends from the S-domain into the C-domain crossing the main P1/P4/P5 coaxial stem (Fig. 2 and Supplementary Fig. 4). The 3' CCA end of the tRNA enters a tunnel formed by P6/P15/P16/P17 and base pairs with nucleotides in the L15 region (Fig. 2 and Supplementary Figs 4 and 5), an interaction recognized previously<sup>23</sup>. The 5' end of the tRNA indicates the location of the active site, which is close to the region where P4, P5 and CR-IV intersect. The protein component is also adjacent to the 5' end of the

<sup>1</sup>Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, USA. <sup>†</sup>Present addresses: Departamento de Bioquímica y Biología Estructural, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Ciudad Universitaria, Apartado Postal 70-243, México 04510, México (A.T.-L.); Abbott Laboratories, Abbott Park, Illinois 60064-6400, USA (K.K.S.).





**Figure 1 | Crystal structure of the *T. maritima* RNase P holoenzyme in complex with tRNA.** **a**, Structure of bacterial RNase P, composed of a large RNA subunit (338 nucleotides, ~110 kDa) and a small protein component (117 amino acids, ~14.3 kDa), in complex with tRNA (76 nucleotides, ~26 kDa). The RNA component serves as the primary biocatalyst in the reaction and contains two domains, termed the catalytic (C, blue) and specificity (S, light blue) domains. The RNase P protein (green) binds the 5' leader region of the pre-tRNA substrate and assists in product release. Transfer RNA (tRNA<sup>Phe</sup>) (red) makes multiple interactions with the P RNA (see Fig. 2 and Supplementary Fig. 1 for details). Regions in grey denote additional RNA nucleotides required for crystallization. **b**, Alternative view of the RNase P–tRNA complex, identifying the tRNA recognition regions: the 5' end where catalysis occurs, the 3' CCA end, and the highly conserved TΨC and D loop regions. **c**, View of the 4.1 Å experimental electron density map centred on the 5' end of tRNA. The map is represented as a dark grey mesh, contoured at 1.4 r.m.s.d.



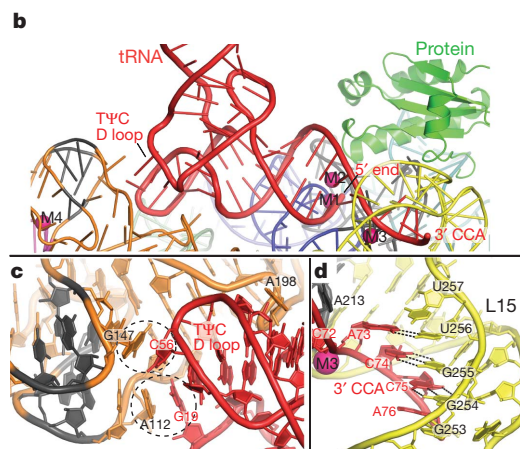
**Figure 2 | tRNA recognition by RNase P is mediated by RNA–RNA interactions.** **a**, Schematic of the P RNA secondary structure mapping the tRNA–P RNA contacts observed in the crystal structure. The tRNA nucleotides (1•72, 2, 3, 64 and 65) and regions (5', 3', TΨC loop, D loop and acceptor) involved in direct interactions are shown in red. Intermolecular base pairs form between the 3' end of tRNA (DCCA) and loop 15 (L15), where D is the discriminator nucleotide that serves as an identity element in tRNA biogenesis. P RNA nucleotides that are universally conserved (black, uppercase), conserved among all bacteria (grey, uppercase), or highly conserved in bacteria (black, lowercase) are identified. Metal ions are shown as filled pink circles, and denote the location of the active site (M1, M2), and other structurally important regions (M3, M4). Single and double dashes in red represent minor groove and base stacking interactions, respectively. All identified tRNA–P RNA contacts are within 4 Å. The crystallized *T. maritima* P RNA consists of eighteen paired helices (P), five universally conserved regions (CR-I to CR-V) (black), two junctions containing conserved nucleotides in bacteria (dark grey), several loop (L) regions, and an engineered tetraloop region (T, light grey). The coaxial P1/P4/P5 stem is shown in blue, P2/P3 stems in cyan, P6/P15/P16 and L15/L17 in

tRNA, but does not contact it. The protein contacts include the CR-IV and CR-V regions, the P15 stem, and the P2/P3 helix interface (Fig. 3 and Supplementary Figs 1 and 6). The pre-tRNA leader makes extensive contacts with the protein, but few with the P RNA.

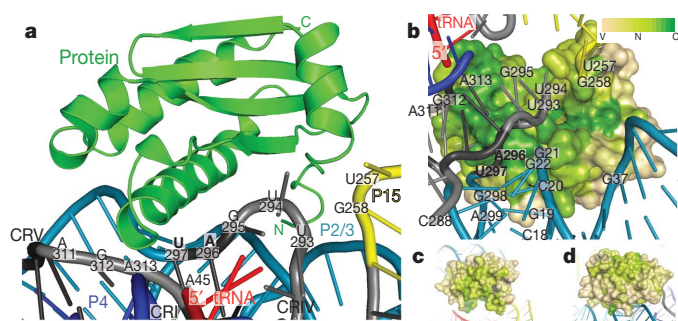
The components of the RNase P holoenzyme are largely unchanged when bound to tRNA (Supplementary Fig. 7). A comparison between *T. maritima* P RNA alone<sup>17</sup> and in the complex reveals an overall similar fold (backbone normalized root mean square deviation (r.m.s.d.) ~1.1 Å) with a small change in the relative orientation of the two domains (Supplementary Fig. 7). The only major change in the P RNA structure occurs in the vicinity of the P15–P17 stems (Supplementary Figs 8 and 9). A few additional residues at the amino terminus were clear and follow a similar path to the *B. subtilis* protein<sup>24</sup> (Supplementary Fig. 10); no changes in the structure of the protein component were detected. The structures of yeast and *T. maritima* tRNA<sup>Phe</sup> show remarkable resemblance (backbone normalized r.m.s.d. for acceptor stem ~0.8 Å) (Supplementary Fig. 11). Further, a comparison with previous models reveals an excellent agreement with the predicted secondary structure<sup>25</sup> and a good agreement with the models of the complex<sup>19,26,27</sup> (Supplementary Fig. 12).

### tRNA recognition

The observed RNA–RNA interactions involved in substrate recognition agree with previous biochemical studies<sup>5,23,28</sup> and include (1) stacking between bases in the tRNA TΨC and D loops and the P RNA S-domain, (2) an A-minor interaction at the acceptor stem, and (3) the formation of canonical base pairs at the 3' end of tRNA (Fig. 2 and Supplementary Fig. 1). The first interaction identifies the TΨC loop as a key element in recognition. Both the tRNA D and TΨC loops have unstacked bases (G19 and C56) that interact with unstacked bases in



yellow, P7 and P10/P11/P12 in orange, P8/P9 in light green, and P13/P14 in pink (see Supplementary Fig. 1 for additional details). **b**, Recognition of tRNA by the P RNA of RNase P. The acceptor stem of tRNA (red) docks onto the P RNA (coloured as in **a**) making a series of interactions, including base stacking in the TΨC/D loops of tRNA and the S-domain, an A-minor interaction, and base pairing, ribose zipper and stacking interactions between the 5' and 3' ends of tRNA and the C-domain. The protein (green) makes no direct contacts with mature tRNA. Critical metal ions (M1–M4) identified are shown as magenta spheres. **c**, tRNA recognition by the S-domain. Two universally conserved P RNA regions (CR-II and III, dark grey) facilitate base stacking interactions with unstacked bases in the structurally conserved TΨC and D loops of tRNA. Dashed circles highlight this stacking interaction between P RNA residues A112, G147 and tRNA residues G19, C56. A conserved P RNA adenosine (A198) stacks into the minor groove of the acceptor tRNA stem. **d**, Recognition of the tRNA 3' CCA by the C-domain. Intermolecular base pairs form between the 3' tRNA (ACC) and the L15 (GGU) loop of P RNA. This interaction is stabilized by a structural metal (M3, magenta sphere) and a L15 ribose zipper conformation.



**Figure 3 | Protein–RNA contacts within the RNase P holoenzyme.** **a**, The protein sits on the P RNA surface formed by conserved regions I, IV and V. The protein (green, shown as ribbons) additionally contacts the L15/P15 junction and the P2/3 helices (P RNA as coloured in Fig. 2). Labelled P RNA nucleotides make protein contacts (within 4 Å) and include: A45 in CR-I, U257 and G258 in the L15/P15 junction, U293, U294, G295, A296 and U297 in CR-IV, and A311, G312 and A313 in CR-V. Bold nucleotides are universally conserved. **b**, Surface representation of the protein coloured by sequence conservation (variable (V), tan; neutral (N), light green; conserved (C), green). A highly conserved patch in the protein extends from the vicinity of the 5' end of the tRNA, and interacts with P RNA conserved regions IV (U293–U297) and V (A311–A313). Other P RNA nucleotides that make protein contacts include: the P2 helix (C18–G22, G298–A299), the P3 helix (G37) and the L15/P15 junction (U257–G258). Four hundred and ninety bacterial RNase P proteins were included in the analysis of the sequence conservation using the ConSurf server<sup>48</sup>. Panels **c** and **d** show different orientations to emphasize that high sequence conservation is concentrated in the region of the protein that faces the conserved regions of the P RNA. Neutral or slightly conserved regions shown in these two orientations correspond to a patch that interacts with the leader.

the P RNA (A112 and G147), forming G19–A112 and C56–G147 stacks in the complex. The second major interaction involves a highly conserved unstacked adenosine (A198) in the P11 stem entering the minor groove of the tRNA acceptor stem. These interactions facilitate shape complementarity and help explain the central role of the S-domain in recognition. The third major interaction involves intermolecular base pairing between the tRNA 3' DCCA motif and the L15 loop. This interaction is probably conserved in all bacterial and most archaeal RNase P, but not in organisms where CCA is added post-transcriptionally<sup>1</sup>. The fourth to last nucleotide, A73, forms a Watson–Crick base pair with nucleotide U256. C74 and C75 form Watson–Crick base pairs with G255 and G254, while the terminal A76 forms a weak interaction with G253. To accommodate these intermolecular base pairs, the two strands of L15 fold into a ribose zipper. In addition, a structural metal ion (M3) (Fig. 2 and Supplementary Fig. 13) binds adjacent to this P RNA–tRNA region and is likely to correspond to a metal ion identified biochemically<sup>7</sup>. In the complex, the 3' end of the tRNA separates from the 5' end and enters a wide opening formed by P6/P15/P16/P17 (Figs 1 and 2 and Supplementary Figs 4 and 5). This opening is ~20 Å in diameter, can easily accommodate a single-stranded RNA molecule, and is created when the P6/L17 pseudoknot forms (Supplementary Figs 1 and 5).

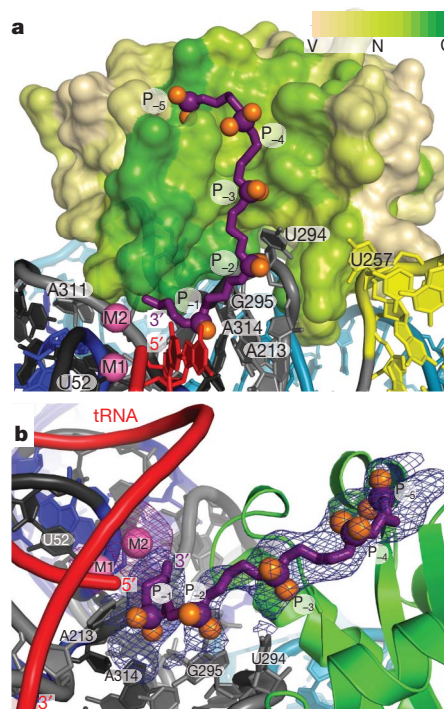
### Protein–RNA interactions

The bacterial RNase P protein structure is highly conserved, but has little or no sequence or structural similarity with the protein components of archaea or eukarya<sup>29</sup>. In the complex, the protein is near the 5' end of tRNA, but is too far (over 6 Å) to make direct contacts. The protein sits between the P15 and P3 stems (Fig. 3 and Supplementary Fig. 6), and also contacts the CR-IV and CR-V loop regions of P RNA. Comparison of bacterial sequences shows that the protein has a large, contiguous area with high sequence conservation (Fig. 3 and Supplementary Fig. 6) including important residues identified previously<sup>11,30,31</sup>. The conserved area extends in an arch along the surface of the protein, starting from a point close to the 5' end of the tRNA and faces the universally conserved modules.

To investigate the interactions with the leader, crystals were soaked with a short oligoribonucleotide in the presence and absence of  $\text{Sm}^{3+}$ . Fourier difference maps to 4.2 Å show five phosphates of the leader along the conserved surface of the protein (Fig. 4 and Supplementary Fig. 6), but the position of the nucleobases was ambiguous. The structure shows that the leader contacts residues Phe 17, Phe 21, Lys 51, Arg 52 and Lys 90 and probably interacts with Gln 28, Lys 56 and Arg 89, in agreement with biochemical results<sup>6,9,12,30,31</sup>. The 3' end of the leader is located adjacent to the 5' end of the tRNA and near two conserved residues (Arg 52 and Lys 56). A metal ion is present in between the leader 3' and the 5' end of mature tRNA (Figs 4 and 5 and Supplementary Fig. 14), but is too distant (>4 Å) to ligate protein residues directly. Leader nucleotides –1 to –3 are poised to interact with nucleotides A213, U294, G295 and A314 of P RNA (Fig. 4 and Supplementary Figs 1 and 6). These results indicate that the major role of the protein component is to interact with the leader to align the pre-tRNA in the complex, as observed previously<sup>9,11,31</sup>.

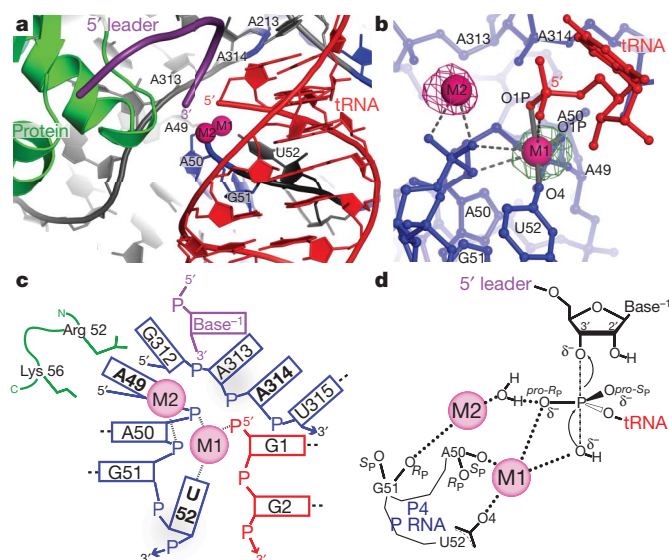
### Active site

The location of the active site is inferred from the 5' end of mature tRNA (Fig. 5 and Supplementary Fig. 15). The phosphate backbone of tRNA nucleotides (+1 to +3) sits on the major groove of the P4 stem (near A50, G51 and U52), and places the tRNA 5' end next to the P4



**Figure 4 | Pre-tRNA leader–protein interactions in the RNase P holoenzyme.** **a**, Surface representation of the protein coloured by sequence conservation as in Fig. 3. The pre-tRNA 5' leader (purple, with purple and orange spheres for the phosphorous and non-bridging oxygens, respectively) was modelled as a polyphosphate chain with five phosphates ( $P_{-1}$  to  $P_{-5}$ ). The leader follows a highly conserved patch in the protein extending from the 5' end of the mature tRNA (red) and away from the P RNA. The addition of a 5' leader with metal ( $\text{Sm}^{3+}$ ) reveals a second metal ion (M2). **b**, Alternative view of the pre-tRNA leader–protein interaction. Each phosphate position ( $P_{-1}$  through  $P_{-5}$ ) was visible in a 4.2 Å difference Fourier map ( $mF_o - DF_c$ ) calculated from crystals where only the leader was soaked into the crystals (blue mesh, 3 r.m.s.d. contour levels). A second 4.2 Å difference Fourier map ( $mF_o - DF_c$ ) calculated from crystals where the leader and  $\text{Sm}^{3+}$  metal were soaked into the crystals shows clearly the position of the second metal ion (magenta mesh, 3.5 r.m.s.d. contour level). P RNA residues poised to make contacts are labelled. Nucleotide U52 serves as a reference point in **a** and **b** and does not interact with the 5' leader oligonucleotide.





**Figure 5 | Structure of the RNase P active site environment.** **a**, The active site is inferred from the location of the mature 5' end of tRNA. The diagram shows the position of the mature tRNA (red), the leader (purple), the protein component (green) and the P RNA (blue and grey). A group of conserved P RNA nucleotides (A49–U52, A213, A313 and A314) form part of the active site. Two metal ions (magenta spheres) are found in the active site. **b**, The two active site metal ions (M1 and M2) are within 4 Å of the 5' phosphate of tRNA and the M1–M2 metal–metal distance is ~4.8 Å. The M1 metal makes contacts ( $\leq 2.1$  Å, solid grey bonds, labelled) with tRNA (G1 O1P) and P RNA (A50 O1P and U52 O4) oxygens. Other possible ligands within 3.5 Å of M1 or M2 are represented by dashed grey lines (Supplementary Table 5). The figure shows two isomorphous difference Fourier ( $mF_o - DF_c$ ) maps. The green mesh corresponds to a  $\text{Eu}^{3+}$  soak in the absence of leader and is contoured at the 9.5 r.m.s.d. level. The magenta mesh corresponds to a  $\text{Sm}^{3+}$  and 5' leader soak and is contoured at the 5.5 r.m.s.d. level. The second metal is clearly visible only when the leader is present. **c**, Schematic diagram of the interactions around the active site. The diagram shows all residues within 8 Å of the 5' phosphorus atom of tRNA. Short dashed lines represent metal ligand distances within 2.2 Å and longer dashed lines represent nucleotides which form canonical base pairs. Nucleotides in bold are universally conserved in P RNA. The P RNA, tRNA, 5' leader, and protein side chains are shown in blue, red, purple and green, respectively. **d**, Proposed reaction mechanism for the endonucleolytic cleavage of pre-tRNA by RNase P based on the structure of the enzyme–product (E–P) complex and previous mechanistic studies<sup>49,50</sup>. The M1 metal distance to the 5' phosphate ligands (Supplementary Table 5) in the E–P complex is consistent with the proposed enzyme–substrate (E–S) transition state. In this proposed reaction scheme, M1 is ~180° from the apical O3' position and activates a hydroxyl nucleophile for an in-line nucleophilic displacement, creating a new bond and displacing the 3' scissile phosphate oxygen. As RNase P proceeds through an  $\text{S}_{\text{N}}2$  reaction pathway, the stereochemistry around the phosphorus atom undergoes a net inversion of configuration. If the *pro-R\_p* (O2P) oxygen coordinates metal in the E–S complex during catalysis, as previously observed<sup>49,50</sup>, this would subsequently allow for the *pro-S\_p* (O1P) oxygen to coordinate metal in the E–P complex, as observed in the crystal structure. Product release could be facilitated by a metal (M2) coordinated water, which would enable proton transfer to the 3' scissile oxygen. The exact active site geometry and identity of other metal ligands in an E–S complex has yet to be established.

phosphate backbone and nucleotides A313 and A314 (Fig. 5 and Supplementary Fig. 15). The universally conserved U52 nucleotide is unstacked from the P4 stem and faces the tRNA 5' end. In addition, the tRNA 1•72 base pair is stabilized by an adenosine stack with A213, a nucleotide conserved in all bacteria.

A metal ion (M1), putatively magnesium, is found trapped between the tRNA 5' end, the A50 and G51 phosphates, and the O4 oxygen of the universal U52 nucleotide and was confirmed using crystals soaked with  $\text{Sm}^{3+}$  and  $\text{Eu}^{3+}$  (Supplementary Figs 14 and 15). Putative M1 metal contacts include the A50 non-bridging phosphoryl oxygen, the

O4 oxygen of the U52 nucleobase, and the O1P oxygen at the 5' end of tRNA. Other metal–ligand interactions may include: the backbone of A50, the phosphoryl oxygen of G51, and the 5' end of tRNA (Supplementary Table 5). Many of these oxygen ligands have been implicated in metal coordination and catalysis<sup>32–34</sup>. The M1 site may also coincide with a site (M6) observed in the structure of *B. stearothermophilus* P RNA<sup>35</sup>. The structure of the complex suggests that M1 participates in catalysis by directly binding P RNA and the 5' phosphate of tRNA.

A second metal (M2) was located in experiments where the leader was soaked in the presence of  $\text{Sm}^{3+}$ . The M2 metal is in close proximity to the phosphoryl oxygens of G51, the O3' of the leader, and the 5' end of tRNA (Supplementary Table 5). The two metals observed in crystals soaked with the leader and  $\text{Sm}^{3+}$  are ~4.8 Å apart (Fig. 5 and Supplementary Fig. 15). The structures indicate that the active site includes at least two metal ions upon complex formation with pre-tRNA. Due to its location, the M2 metal ion could make additional contacts with both the tRNA and the P RNA during catalysis.

The structures of the active site of the complex and the aporibozyme structures are similar (Supplementary Figs 7, 16 and 17), including the presence of a metal ion next to the P4 helix<sup>35</sup>. With the exception of the U52 nucleobase (Supplementary Fig. 16), no large changes are observed in the active site region. A fully occupied M2 site is observed only in the presence of leader, suggesting that a local metal-dependent conformation change may occur, as previously reported<sup>6</sup>. The structure also reveals that the tRNA 5' and 3' ends splay and separate to interact with the P RNA (Supplementary Fig. 11), confirming the need for movement of the tRNA ends<sup>36,37</sup>. Although accommodating the upstream RNA leader probably requires local protein and P RNA structural changes, the location of the active site is not significantly altered and is largely pre-assembled.

## Mechanistic implications

RNase P can cleave a variety of substrates<sup>1,10,38</sup>, but pre-tRNA is the only one that is common among all organisms. To decipher its function, it is important to understand two different aspects of pre-tRNA processing by RNase P: substrate specificity and the chemical mechanism of cleavage.

tRNA recognition by RNase P involves the highly conserved tRNA TΨC and D loops and the CR-II and CR-III in the S-domain of P RNA. Thus, regions with high sequence and structure conservation are involved in specific tertiary interactions, suggesting a universal mode of recognition among all RNase P. The presence of unpaired nucleotides next to the cleavage site is also an important feature for pre-tRNA recognition, although it is unclear whether this is a universal feature of all natural substrates<sup>1</sup>. Finally, pre-tRNA is usually processed to form a 7-base-pair-long acceptor stem. An additional role of the interactions between CR-II and CR-III and tRNA may be to serve as a 'ruler' that ensures that the correct lengths are processed, although there is some flexibility as tRNAs with acceptor stems 8 base pairs long can be processed<sup>39</sup>. The interaction with the 3' CCA end is also a key recognition feature, but may not be necessarily an RNA–RNA interaction in higher organisms. The L15 loop of P RNA is not found in eukarya or some archaea<sup>40</sup> and its function may be replaced by additional protein(s), suggesting that 3' CCA intermolecular base pairing is not a universal interaction.

The second important aspect of RNase P function is the chemical mechanism of cleavage. Hydrolysis of a phosphodiester bond generates the mature 5' RNA product. Whereas it is not possible to propose a complete mechanism from a structure at this resolution, the RNase P–tRNA structures, together with extensive biochemical information, help identify the major active site components. The structure indicates that at least two distinct metals play a direct role. It is possible to propose a transition state model (Fig. 5d) where the M1 metal directly positions the scissile phosphate oxygens of the substrate and enables a hydroxyl ion to perform an  $\text{S}_{\text{N}}2$ -type nucleophilic substitution. In this scenario,



the M2 metal ion stabilizes the transition state and mediates proton transfer to the 3' scissile oxygen during product release, as proposed previously<sup>7</sup>. Other universally conserved nucleotides in the vicinity seem to have a structural role in forming the correct structure and are not directly involved in catalysis, consistent with proposals that sequence conservation is largely the result of strong structural constraints<sup>19</sup>. Hence, the RNase P–tRNA complex shows how the P RNA structure can serve as a scaffold to bind and orient metals and substrate properly. It seems that RNase P uses a two-metal ion catalytic mechanism, similar to other mechanisms proposed based on other large ribozyme structures<sup>41,42</sup> and originally put forth as a general mechanism for many ribozymes<sup>43</sup>.

The structural studies of the holoenzyme–tRNA complex help to show that all RNase P ribozymes share a common, RNA-based mechanism of RNA cleavage and recognition that involves two universally conserved structural modules. Adaptation through the addition of protein increases RNase P functionality by positioning accurately the 5' leader pre-tRNA substrate and by contacting conserved regions of the P RNA structure. The unique tertiary fold of the P RNA uses shape complementarity, specific RNA–RNA contacts, and intermolecular base pairing to recognize its substrate efficiently. Within this tertiary fold, the universally conserved regions are crucial to form the active site scaffold and to create regions involved in tRNA recognition. In addition, both P RNA and the pre-tRNA help to coordinate two catalytically important metal ions essential for the putative mechanism of pre-tRNA cleavage. The RNase P–tRNA complex offers a glimpse into the transition from an ancient, RNA-based world to the present, protein-catalyst dominated world and affirms that RNA molecules can display comparable versatility and complexity.

## METHODS SUMMARY

**Crystallization.** Preparation, purification and folding of *T. maritima* RNase P and tRNA<sup>Phe</sup> have been described<sup>8,17,44</sup>. For crystallization, the components were mixed in a 1:1.1:1 (P RNA:pre-tRNA:protein) molar ratio to a concentration of 45  $\mu$ M. The mixture was heated to 94 °C (2 min), cooled to 4 °C (2 min), and after the addition of MgCl<sub>2</sub> to a final 10 mM concentration, further incubated at 50 °C (10 min) and 37 °C (40 min). Crystals were obtained by mixing 1  $\mu$ l of complex with 1  $\mu$ l of reservoir solution (1.8 M Li<sub>2</sub>SO<sub>4</sub>, 50 mM sodium cacodylate (pH 6.0)) and equilibrated by vapour diffusion at 30 °C. Crystals were cryo-protected using reservoir solution containing 15% xylitol.

**Data collection and structure determination.** Diffraction data were collected at 100 K at the LS-CAT sector at the APS. Complete native and Ta<sub>6</sub>Br<sub>12</sub>, SmCl<sub>3</sub>, EuCl<sub>3</sub> and iridium hexammine (Ir(NH<sub>3</sub>)<sub>6</sub>)<sup>3+</sup> derivatives were collected. A weak Molecular Replacement<sup>45</sup> solution using a trimmed model of the tRNA–P RNA complex<sup>19</sup> located the Ta<sub>6</sub>Br<sub>12</sub> cluster. Multi-wavelength anomalous dispersion (MAD) phases<sup>46</sup> from the cluster extended to ~6 Å, with the map showing a clear envelope. These phases were used to locate the other heavy atoms that were used to calculate a 4.1 Å MIRAS map. To locate the pre-tRNA leader, crystals were soaked with a *T. maritima* 5' tRNA 7-nucleotide leader sequence (final concentration 0.2 mM), with and without 14 mM SmCl<sub>3</sub>. Difference maps allowed the placement of five pre-tRNA nucleotides and the unambiguous identification of a second active site metal. The experimental electron density map was of excellent quality and allowed model building of nearly all RNA phosphate and nucleobase positions and accurate placing of the protein. Model building was guided by the known structures<sup>14,17,21</sup>. Final *R*<sub>work</sub> and *R*<sub>free</sub> are 24.9% and 27.0%, respectively, with r.m.s.d. of 0.007 Å and 1.24° for bonds and angles. Figures were made with PyMOL<sup>47</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 4 June; accepted 17 September 2010.

Published online 14 November 2010.

- Hartmann, R. K., Gossringer, M., Spath, B., Fischer, S. & Marchfelder, A. The making of tRNAs and more – RNase P and tRNase Z. *Prog. Mol. Biol. Transl. Sci.* **85**, 319–368 (2009).
- Kazantsev, A. V. & Pace, N. R. Bacterial RNase P: a new view of an ancient enzyme. *Nature Rev. Microbiol.* **4**, 729–740 (2006).
- Liu, F. & Altman, S. *Protein Reviews Volume 10: Ribonuclease P* (Springer, 2010).
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).

- Pan, T., Loria, A. & Zhong, K. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. *Proc. Natl Acad. Sci. USA* **92**, 12510–12514 (1995).
- Hsieh, J. *et al.* A divalent cation stabilizes the active conformation of the B. subtilis RNase P pre-tRNA complex: a role for an inner-sphere metal ion in RNase P. *J. Mol. Biol.* **400**, 38–51 (2010).
- Kirsebom, L. A. in *Protein Reviews Volume 10: RNase P* (eds Liu, F. & Altman, S.) Ch. 7, 113–134 (Springer, 2010).
- Buck, A. H., Dalby, A. B., Poole, A. W., Kazantsev, A. V. & Pace, N. R. Protein activation of a ribozyme: the role of bacterial RNase P protein. *EMBO J.* **24**, 3360–3368 (2005).
- Kurz, J. C., Niranjanakumari, S. & Fierke, C. A. Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNA<sup>Asp</sup>. *Biochemistry* **37**, 2393–2400 (1998).
- Peck-Miller, K. A. & Altman, S. Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J. Mol. Biol.* **221**, 1–5 (1991).
- Koutmou, K. S. *et al.* Protein-precursor tRNA contact leads to sequence-specific recognition of 5' leaders by bacterial ribonuclease P. *J. Mol. Biol.* **396**, 195–208 (2010).
- Sun, L., Campbell, F. E., Zahler, N. H. & Harris, M. E. Evidence that substrate-specific effects of C5 protein lead to uniformity in binding and catalysis by RNase P. *EMBO J.* **25**, 3998–4007 (2006).
- Reich, C., Olsen, G. J., Pace, B. & Pace, N. R. Role of the protein moiety of ribonuclease P, a ribonucleoprotein enzyme. *Science* **239**, 178–181 (1988).
- Kazantsev, A. V. *et al.* Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl Acad. Sci. USA* **102**, 13392–13397 (2005).
- Krasilnikov, A. S., Xiao, Y., Pan, T. & Mondragón, A. Basis for structural diversity in homologous RNAs. *Science* **306**, 104–107 (2004).
- Krasilnikov, A. S., Yang, X., Pan, T. & Mondragón, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760–764 (2003).
- Torres-Larios, A., Swinger, K. K., Krasilnikov, A. S., Pan, T. & Mondragón, A. Crystal structure of the RNA component of bacterial ribonuclease P. *Nature* **437**, 584–587 (2005).
- Chen, J.-L. & Pace, N. R. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3**, 557–560 (1997).
- Torres-Larios, A., Swinger, K. K., Pan, T. & Mondragón, A. Structure of ribonuclease P—a universal ribozyme. *Curr. Opin. Struct. Biol.* **16**, 327–335 (2006).
- Ferré-d'Amaré, A. R., Zhou, K. & Doudna, J. A. A general module for RNA crystallization. *J. Mol. Biol.* **279**, 621–631 (1998).
- Kazantsev, A. V. *et al.* High-resolution structure of RNase P protein from *Thermotoga maritima*. *Proc. Natl Acad. Sci. USA* **100**, 7497–7502 (2003).
- Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**, 1091–1105 (2000).
- Kirsebom, L. A. & Svard, S. G. Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* **13**, 4870–4876 (1994).
- Stams, T., Niranjanakumari, S., Fierke, C. A. & Christianson, D. W. Ribonuclease P protein structure: evolutionary origins in the translational apparatus. *Science* **280**, 752–755 (1998).
- Massire, C., Jaeger, L. & Westhof, E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279**, 773–793 (1998).
- Buck, A. H., Kazantsev, A. V., Dalby, A. B. & Pace, N. R. Structural perspective on the activation of RNase P RNA by protein. *Nature Struct. Mol. Biol.* **12**, 958–964 (2005).
- Tsai, H. Y., Masquida, B., Biswas, R., Westhof, E. & Gopalan, V. Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. *J. Mol. Biol.* **325**, 661–675 (2003).
- LaGrande, T. E., Huttenhofer, A., Noller, H. F. & Pace, N. R. Phylogenetic comparative chemical footprint analysis of the interaction between ribonuclease P RNA and tRNA. *EMBO J.* **13**, 3945–3952 (1994).
- Mondragón, A. in *Protein Reviews Volume 10: Ribonuclease P* (eds Liu, F. & Altman, S.) Ch. 4, 63–78 (Springer, 2010).
- Biswas, R., Ledman, D. W., Fox, R. O., Altman, S. & Gopalan, V. Mapping RNA-protein interactions in ribonuclease P from *Escherichia coli* using disulfide-linked EDTA-Fe. *J. Mol. Biol.* **296**, 19–31 (2000).
- Niranjanakumari, S., Stams, T., Crary, S. M., Christianson, D. W. & Fierke, C. A. Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA. *Proc. Natl Acad. Sci. USA* **95**, 15212–15217 (1998).
- Christian, E. L., Kaye, N. M. & Harris, M. E. Helix P4 is a divalent metal ion binding site in the conserved core of the ribonuclease P ribozyme. *RNA* **6**, 511–519 (2000).
- Crary, S. M., Kurz, J. C. & Fierke, C. A. Specific phosphorothioate substitutions probe the active site of *Bacillus subtilis* ribonuclease P. *RNA* **8**, 933–947 (2002).
- Christian, E. L., Smith, K. M., Perera, N. & Harris, M. E. The P4 metal binding site in RNase P RNA affects active site metal affinity through substrate positioning. *RNA* **12**, 1463–1467 (2006).
- Kazantsev, A. V., Krivenko, A. A. & Pace, N. R. Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA* **15**, 266–276 (2009).
- Pomeranz Krummel, D. A., Kent, O., MacMillan, A. M. & Altman, S. Evidence for helical unwinding of an RNA substrate by the RNA enzyme RNase P: use of an interstrand disulfide crosslink in substrate. *J. Mol. Biol.* **295**, 1113–1118 (2000).
- Gaur, R. K., Hanne, A., Conrad, F., Kahle, D. & Krupp, G. Differences in the interaction of *Escherichia coli* RNase P RNA with tRNAs containing a short or a long extra arm. *RNA* **2**, 674–681 (1996).
- Forster, A. C. & Altman, S. Similar cage-shaped structures for the RNA components of all ribonuclease P and ribonuclease MRP enzymes. *Cell* **62**, 407–409 (1990).

39. Burkard, U. & Soll, D. The unusually long amino acid acceptor stem of *Escherichia coli* selenocysteine tRNA results from abnormal cleavage by RNase P. *Nucleic Acids Res.* **16**, 11617–11624 (1988).
40. Walker, S. C. & Engelke, D. R. Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.* **41**, 77–102 (2006).
41. Stahley, M. R. & Strobel, S. A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587–1590 (2005).
42. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).
43. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90**, 6498–6502 (1993).
44. Krivenko, A. A., Kazantsev, A. V., Adamidi, C., Harrington, D. J. & Pace, N. R. Expression, purification, crystallization and preliminary diffraction analysis of RNase P protein from *Thermotoga maritima*. *Acta Crystallogr. D* **58**, 1234–1236 (2002).
45. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
46. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
47. DeLano, W. L. The PyMOL molecular graphics system. (<http://www.pymol.org>) (2002).
48. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).
49. Chen, Y., Li, X. & Gegenheimer, P. Ribonuclease P catalysis requires  $Mg^{2+}$  coordinated to the *pro-Rp* oxygen of the scissile bond. *Biochemistry* **36**, 2425–2438 (1997).
50. Warnecke, J. M. *et al.* Ribonuclease P (RNase P) RNA is converted to a  $Cd^{2+}$ -ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase P cleavage site. *Proc. Natl Acad. Sci. USA* **93**, 8924–8928 (1996).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Sontheimer and O. Uhlenbeck for comments and suggestions, N. Pace for the gift of the *T. maritima* RNase P protein plasmid, Obiter Research, A. Davis and M. E. Duban for advice and preparation of iridium hexammine, and A. Samelson for discussions and assistance. In addition, we are grateful for data collection assistance from S. Anderson, Z. Wawrzak, and staff at LS-CAT. Research was supported by the NIH. N.J.R. is an NRSA postdoctoral fellow.

**Author Contributions** A.M. directed the work. A.T.-L. and A.M. conceived the project. All authors performed and designed experiments. N.J.R. obtained crystallographic data. N.J.R. and A.M. analysed the crystallographic data. N.J.R. and A.M. wrote the paper with contributions from all authors.

**Author Information** Coordinates for the RNase P holoenzyme–tRNA complex, with and without 5' tRNA leader sequence, have been deposited into the RCSB Protein Data Bank (accession code 3OKB and 3OK7, respectively). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.M. ([a-mondragon@northwestern.edu](mailto:a-mondragon@northwestern.edu)).

## METHODS

**Preparation of the *T. maritima* RNase P holoenzyme–tRNA<sup>Phe</sup> ternary complex.** RNA transcriptions were performed *in vitro* using purified His<sub>6</sub>-tagged T7 RNA polymerase using standard protocols<sup>51</sup>. Sequences from the *T. maritima* RNase P RNA and tRNA<sup>Phe</sup> genes were inserted into a pUC19 vector at FokI and BsmAI restriction sites, respectively, allowing for run-off transcription of the DNA plasmid after digestion with the appropriate restriction enzyme (NEB). Constructions of modified RNA molecules with either mutations or additions were performed using a QuikChange mutagenesis kit (Stratagene). RNA samples were purified by 6% denaturing polyacrylamide gel electrophoresis (PAGE), identified by ultraviolet absorbance, recovered by diffusion into 50 mM potassium acetate (pH 7) and 0.2 M potassium chloride, and precipitated with ethanol. tRNA was further purified by anion exchange (MonoQ (5/50 GL)) and gel filtration (HiPrep 26/60, Sephacryl S-200) chromatography (GE Health Sciences). Over-expression and purification of the RNase P protein from *T. maritima* was performed as described previously<sup>44</sup>.

To form the RNase P holoenzyme–tRNA complex, unfolded P RNA, unfolded tRNA and P protein molecules were mixed at a 1:1.1:1 molar ratio in 66 mM HEPES, 33 mM Tris (pH 7.4), 0.1 mM EDTA (1× THE) and 100 mM CH<sub>3</sub>COONH<sub>4</sub> (Ref. 8). The ternary mix, at a final concentration of 45 μM, was incubated at 94 °C for 2 min and then cooled to 4 °C over 2 min. After addition of MgCl<sub>2</sub> to a final 10 mM concentration, the reaction mixture was incubated at 50 °C for 10 min, followed by incubation at 37 °C for 40 min, and finally cooled to 4 °C over 30 s.

**Rational design of an RNA tertiary module to build a crystal lattice.** To promote formation of a crystal lattice, intermolecular interactions were facilitated by introducing a tertiary structure interaction module. Based on the *T. maritima* RNA sequence and a proposed model of the P RNA–tRNA complex<sup>19</sup>, constructs were designed where a tetraloop was inserted into the P12 loop (L12) of the P RNA and a tetraloop–receptor into the anticodon stem of tRNA (Supplementary Fig. 1). These two RNA regions were chosen as they were deemed to be far from the active site or other regions involved in specific interactions. In addition, the P12 stem of P RNA has a highly variable helix length across all organisms, lacks sequence conservation, and is non-essential or absent in several organisms<sup>40</sup>. The P12 and the anticodon loop of tRNA are not known to form any functional contacts. The length of the anticodon and the P12 stems were systematically varied by single base pair insertions adjacent to the tetraloop and tetraloop receptor module, thus altering the position (~2.7 Å per base pair added) and orientation (~36° per base pair added) of the tetraloop receptor and the tetraloop. Forty two combinations of molecules were screened for crystallization conditions using a sparse matrix approach employing a set of crystallization conditions developed locally. A few combinations of RNA molecules produced crystals, with most of them diffracting poorly. The best crystals were obtained from a construct where the P12 and anti-codon stems were elongated by five and three base pairs respectively. Insertion of two G–U wobble pairs adjacent to the tetraloop–tetraloop receptor module further improved diffraction, and also created a binding site for an iridium hexammine cation.

**Crystallization and data collection.** Crystals were obtained by mixing 1 μl of complex with 1 μl of reservoir solution (1.8 M LiSO<sub>4</sub>, 50 mM sodium cacodylate (pH 6.0)) and equilibrated by vapour diffusion hanging or sitting drops at 30 °C. Gel analysis of washed crystals show that all three components were present (data not shown). Attempts to crystallize the complex in the absence of protein yielded no crystals. Crystals suitable for data collection grew in approximately 3 weeks and were cryo-cooled in liquid nitrogen immediately after transfer to reservoir solution containing 15% xylitol. Crystals of the RNase P holoenzyme–tRNA ternary complex suitable for data collection grew to approximately ~80–300 μM per side/edge, and diffract anisotropically to 3.8 Å in the best direction and ~4.0 Å in other directions. Crystals belong to space group P3<sub>1</sub>21 (*a* = *b* = 169.3 Å, *c* = 185 Å) and contain one molecule per asymmetric unit.

A series of derivatized crystals were also prepared by soaking in heavy metal compounds. Derivatives were prepared by soaking the crystals in mother liquor plus the derivative and incubating for 2–24 h before transferring them to cryoprotectant with the derivative present and freezing them in liquid nitrogen. Successful derivatizations were obtained by soaking the crystals in the following compounds: 2 mM Ta<sub>6</sub>Br<sub>12</sub>, 14 mM samarium chloride (Sm<sup>3+</sup>), 14 mM europium chloride (Eu<sup>3+</sup>), and 15 mM iridium hexammine (Ir(NH<sub>3</sub>)<sub>6</sub>)<sup>3+</sup>. However, several of the compounds partially precipitated upon addition to the mother liquor solution and hence the final concentration is not known precisely. In addition, crystals with a leader present were obtained by soaking in a 0.2 mM heptamer oligonucleotide (5′-A<sub>7</sub>-G<sub>5</sub>-G<sub>5</sub>-G<sub>4</sub>-C<sub>3</sub>-G<sub>2</sub>-U<sub>1</sub>-3′) (Thermo Fisher) for 4 h with and without 14 mM samarium chloride present. The sequence was chosen by selecting the most common nucleotide in the *T. maritima* tRNA leaders at each position.

All diffraction data were collected at 100 K at the Life Science-Collaborative Access Team (LS-CAT) sector located at the Advance Photon Source (APS) using Rayonix CCD detectors. As the crystals are very radiation sensitive, the data collection range was optimized using the program MOSFLM<sup>52</sup> to collect the most complete native or anomalous data set using the minimal rotation range. Multi-wavelength anomalous dispersion (MAD) data were collected from a tantalum bromide cluster (Ta<sub>6</sub>Br<sub>12</sub>) derivative at three different wavelengths. Single or multiple wavelength anomalous dispersion data were also collected from the samarium chloride (Sm<sup>3+</sup>), europium chloride (Eu<sup>3+</sup>), and iridium hexammine (Ir(NH<sub>3</sub>)<sub>6</sub>)<sup>3+</sup> derivatives. Data were processed with XDS<sup>53</sup> and scaled with SCALA<sup>54</sup>. All other processing was done with programs from the CCP4 suite<sup>54</sup>, except when noted. Data collection statistics for native and derivative data sets are shown in Supplementary Table 2.

In all cases, the diffraction limits of the data were anisotropic. The extent of the anisotropy was determined using the Anisotropy Server<sup>55</sup> and the data were treated in three different ways: (1) without any anisotropy correction; (2) carving the data to the limits suggested by the anisotropy server (3σ cut-off level on amplitudes); and (3) applying an anisotropic correction to the data using the server. For the second case, the integrated data from XDS was carved to the limits suggested by the server and then merged and scaled with SCALA before final processing. In many instances, the phasing and refinement calculations were done separately with the complete and carved data sets and the results compared. Overall, the different ways of treating the data had little effect on the final results, even though the data collection statistics were better for the carved data set (see Supplementary Tables 2 and 3).

**Structure determination and model refinement.** Molecular replacement (MR) studies with the program PHASER<sup>45</sup> using a proposed partial model of the P RNA–tRNA complex<sup>19</sup> gave a weak low resolution (25–8 Å) MR solution (Z-scores: 5.4 and 9.0 for the rotation and translation functions, respectively). Phases calculated from the MR solution were used to locate the position of the three sites in the Ta<sub>6</sub>Br<sub>12</sub> cluster data set. The program SHARP<sup>46</sup> was used to calculate MAD phases using data from three different wavelengths and spherically averaged form factors for the cluster. The solvent-flattened MAD map was of excellent quality but the phases were only good to ~6 Å resolution. The positions of the Eu<sup>3+</sup>, Sm<sup>3+</sup> and (Ir(NH<sub>3</sub>)<sub>6</sub>)<sup>3+</sup> heavy atoms were determined using the cluster phases. The parameters from the cluster and other derivatives could not be refined simultaneously and instead multiple isomorphous replacement with anomalous scattering (MIRAS) phases to 4.1 Å resolution were calculated using data from the single-atom derivatives together with phase information to 6 Å from the cluster data. The SOLOMON<sup>56</sup> solvent-flattened map was very clear (Supplementary Fig. 2) and all three molecules were apparent in the map. The model for the P RNA–tRNA complex<sup>19</sup> fit well in many areas, but the map showed regions where the model needed to be changed, regions that were missing in the model, like the P12 extension and the pseudoknot region, and the position of the protein. The models for the tRNA and P RNA were rebuilt completely using the high resolution model of yeast tRNA<sup>Phe</sup> (ref. 22), *T. maritima* P RNA<sup>17</sup> and *B. stearothermophilus* P RNA<sup>14</sup> as guides. All regions of the RNA molecules were visible in the map and regions that were missing in the original *T. maritima* P RNA model were built. Some minor corrections to the original model were needed, but overall the models for P RNA agree well. The protein density was clear for the backbone, but not for the side chains and hence the high resolution model of the *T. maritima* protein<sup>21</sup> was placed on the experimental electron density map as a rigid body with minimal rebuilding.

Refinement was performed using Refmac5<sup>57</sup> and BUSTER<sup>58</sup>. Owing to the resolution of the data, the models were restrained to enforce good hydrogen bonding distance between Watson–Crick base pairs, planarity between base pairs (both for Watson–Crick and non-Watson–Crick base pairs), and C3′-endo sugar puckering for recognizable secondary structure elements. In addition, during BUSTER refinement the protein was restrained by the high resolution structure of the protein<sup>21</sup>. Model building with Cool<sup>59</sup> was interspersed with either Refmac5 or BUSTER refinement. During rebuilding, missing nucleotides were added as well as some missing residues at the N terminus of the protein. Mg<sup>2+</sup> ions were included at positions that had high density peaks in residual maps and also coincided with heavy atom sites. Other large peaks in the native data set that coincided with phosphate positions in the leader-soaked crystals were modelled as phosphate ions. No individual atomic or group temperature factors were refined, only an overall anisotropic temperature factor. The final stages of the refinement were done with the program BUSTER. The refinement was done both with a carved data set where data outside the anisotropic diffraction limits (3σ cut-off) were excluded and also with a complete data set (isotropic) to the highest resolution limit (Supplementary Table 3). No significant difference was noted in the two refinements and the refinement statistics and electron density maps calculated from either data set were also virtually identical. It seems that the



anisotropic temperature factor correction in the refinement programs adequately modelled the modest anisotropy of the data.

The final model for the P RNA includes nucleotides 1 to 338. Only the phosphate backbone was modelled for nucleotides 39, 241 and 314–317. In addition, 9 nucleotides were inserted between nucleotides 130 and 136 to account for the extension added for crystallization (Supplementary Fig. 1). The final model for the tRNA includes nucleotides 1–76, but only the phosphate backbone was modelled for nucleotides 16, 17 and 20. The crystallization module added eight extra nucleotides incorporated at the end of the anticodon stem (Supplementary Fig. 1). Nearly the entire anticodon stem and anticodon loop were altered to accommodate the tetraloop receptor and altered anticodon loop. The protein model includes residues 6 to 117. The positions of all side chains were ambiguous in the map and were not rebuilt, but kept as much as possible as in the original 1.2 Å model (PDB ID 1NZ0) during refinement. Side chains that collided with the RNA were rebuilt when needed. There are four  $Mg^{2+}$  and two phosphate ions in the model of the complex. The final model to 3.8 Å resolution has an overall  $R_{work}$  of 24.9% and  $R_{free}$  of 27.0% with a root mean square deviation (r.m.s.d.) from target values of 0.007 Å and 1.24° for bonds and angles, respectively. The model in the presence of the leader includes an additional polyphosphate molecule with five phosphates and two  $Mg^{2+}$  ions coinciding with metals ions M1 and M2. A total of five  $Mg^{2+}$  ions were modelled into the complex that contains the 5' polyphosphate leader backbone. The final model to 4.21 Å resolution has an overall  $R_{work}$  of 25.8% and  $R_{free}$  of 26.7% with an r.m.s.d. of 0.007 Å and 1.23° for bonds and angles, respectively (see Supplementary Tables 3 and 4).

Model superpositions were done with programs from the CCP4 suite<sup>54</sup>, lsqman<sup>60</sup> and Coot<sup>59</sup>. Diagrams were made with PYMOL<sup>47</sup>. Coordinates and structure factors have been deposited in the PDB with accession numbers 3OK7 and 3OKB.

**Activity assays of RNase P holoenzyme.** Cleavage assays measuring  $k_{cat}/K_M$  under single turnover conditions were performed on the RNase P and pre-tRNA constructs that gave the best diffracting crystals. The pre-tRNA (with a single nucleotide leader (–1)) which yielded crystals and a control pre-tRNA (containing a *T. maritima* nine nucleotide leader (–9)) were radioactively labelled at their 5' ends. Labelled substrates were purified over a 10% denaturing polyacrylamide gel and identified by <sup>32</sup>P-phosphorimaging. The holoenzyme was folded and cleavage reactions were performed in identical conditions as the folding reaction (1× THE, 10 mM  $MgCl_2$ , 0.1 M  $CH_3COONH_4$ , 37 °C). The enzyme activity of both the modified RNase P which gave crystals and the *T. maritima* wild-type RNase P were tested. The reaction was initiated by mixing

pre-folded RNase P holoenzyme (25, 50 and 100 nM) with pre-folded pre-tRNA substrate (<4 nM), incubated for various times ( $t = 0, 0.25, 1, 4$  and 16 min), and subsequently quenched by adding 9 M urea, 50 mM EDTA. All reaction mixtures were loaded directly on a 15% denaturing polyacrylamide gel which separated the substrate from the product(s). To observe unambiguously the products of the leader (–1) pre-tRNA, thin layer chromatography (TLC) was also performed with polyethyleneimine (PEI)-cellulose coated plates, where the quenched reaction mixture was spotted and run in a 5% acetic acid/100 mM  $NH_4Cl$  solution. The dried gels and the TLC plates were exposed to a phosphorimaging screen and the reaction profile was quantified by a phosphorimager (Fuji Medical) using ImageGauge software. A plot of the percentage of product over time gave the cleavage reaction rate for each concentration. Single turnover conditions assuming a first order reaction follow the equation  $P = P_{\infty}(1 - e^{-k_{obs}t})$ , where  $P$  is the fraction of pre-tRNA cleaved,  $P_{\infty}$  is the fraction of uncleaved pre-tRNA at the end of the reaction, and  $k_{obs}$  is the observed reaction rate constant. By measuring  $k_{obs}$  at different concentrations it is possible to obtain  $K_{cat}/K_M$  assuming Michaelis–Menten kinetics.

51. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).
52. Leslie, A. G. W. in *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**, 27–33 (1992).
53. Kabsch, W. Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* **21**, 67–72 (1988).
54. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
55. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
56. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
57. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
58. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
59. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
60. Kleywegt, G. J. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D* **52**, 842–857 (1996).

# Acid sensing by the *Drosophila* olfactory system

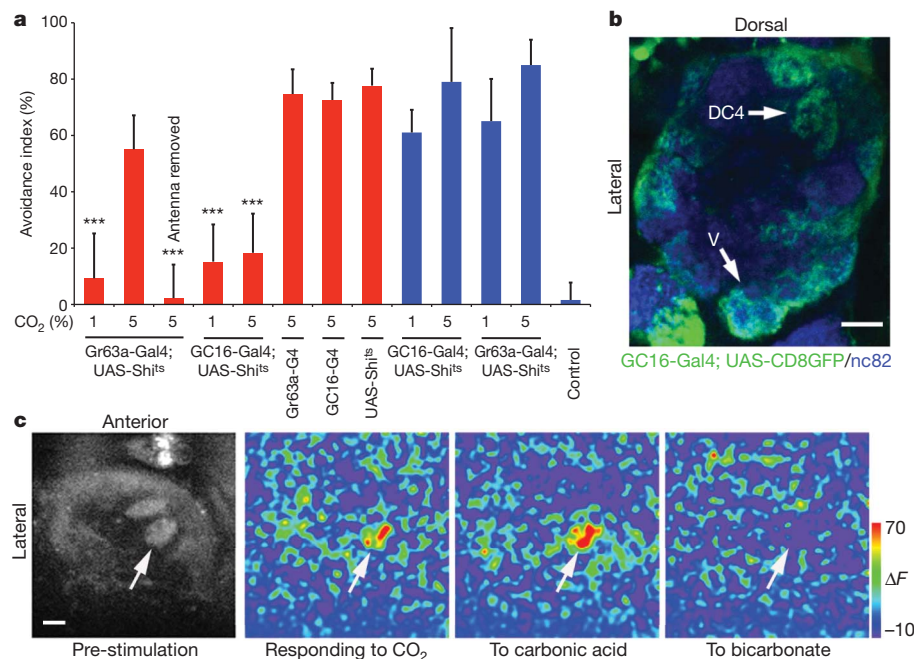
Minrong Ai<sup>1\*</sup>, Soohong Min<sup>1\*</sup>, Yael Grosjean<sup>2†</sup>, Charlotte Leblanc<sup>1†</sup>, Rati Bell<sup>2</sup>, Richard Benton<sup>2</sup> & Greg S. B. Suh<sup>1</sup>

The odour of acids has a distinct quality that is perceived as sharp, pungent and often irritating<sup>1</sup>. How acidity is sensed and translated into an appropriate behavioural response is poorly understood. Here we describe a functionally segregated population of olfactory sensory neurons in the fruitfly, *Drosophila melanogaster*, that are highly selective for acidity. These olfactory sensory neurons express IR64a, a member of the recently identified ionotropic receptor (IR) family of putative olfactory receptors<sup>2</sup>. *In vivo* calcium imaging showed that IR64a+ neurons projecting to the DC4 glomerulus in the antennal lobe are specifically activated by acids. Flies in which the function of IR64a+ neurons or the *IR64a* gene is disrupted had defects in acid-evoked physiological and behavioural responses, but their responses to non-acidic odors remained unaffected. Furthermore, artificial stimulation of IR64a+ neurons elicited avoidance responses. Taken together, these results identify cellular and molecular substrates for acid detection in the *Drosophila* olfactory system and support a labelled-line mode of acidity coding at the periphery.

Many aversive odors activate combinations of olfactory sensory neurons (OSNs)<sup>3,4</sup>, complicating the dissection of the circuits that translate odour recognition into behaviour. By contrast, carbon dioxide (CO<sub>2</sub>), an odorant that is salient for many insect behaviours<sup>5–7</sup>, activates

a single population of dedicated sensory neurons expressing the gustatory receptors GR21a and GR63a (refs 7–9). These neurons are essential for mediating avoidance behaviour of *Drosophila* to CO<sub>2</sub> at concentrations lower than about 2% (refs 7, 8, 10). However, we found that flies in which GR21a/GR63a+ neurons were inactivated still avoided CO<sub>2</sub> at concentrations higher than about 5% (Fig. 1a). Avoidance of high CO<sub>2</sub> concentrations required the antennae (Fig. 1a), indicating that another population of antennal neurons mediates avoidance to high CO<sub>2</sub> concentrations.

To identify these sensory neurons, we performed a functional screen for neurons required for responsiveness to CO<sub>2</sub> by crossing a collection of GAL4 enhancer traps to UAS-*Shibire*<sup>ts</sup> (ref. 11). We isolated a line, GC16-GAL4, that failed to avoid 1% and 5% CO<sub>2</sub> (Fig. 1a). GC16-GAL4 is expressed in OSNs that project to the V glomerulus among others, which is consistent with its defect in avoidance to 1% CO<sub>2</sub> (Fig. 1b and Supplementary Fig. 1). To test whether other glomeruli labelled by GC16-GAL4 besides V are activated by CO<sub>2</sub>, we conducted *in vivo* calcium imaging<sup>12</sup> of the antennal lobe of flies carrying GC16-GAL4 and UAS-GCaMP, a calcium-sensitive green fluorescent protein (GFP)<sup>13</sup>. Using this approach, we identified an additional pair of dorsal glomeruli, termed DC4 (ref. 14), that were activated by about 5% CO<sub>2</sub> (Fig. 1c).



**Figure 1 | Identification of a glomerulus, DC4, activated by the CO<sub>2</sub> metabolite carbonic acid.** **a**, Behavioural testing in a T-maze at permissive (21 °C, blue) and non-permissive (29/34 °C (see Methods), red) temperatures. 'Control' in this and all subsequent figures refers to responses of flies given a choice between two blank tubes. Error bars indicate s.e.m. (n = 6–8). Three

asterisks, P < 0.001 (analysis of variance, Tukey test). **b**, A single optical plane of the antennal lobe illustrates that the DC4 and V glomeruli, among others, are labelled by GC16-GAL4. **c**, *In vivo* calcium imaging of the antennal lobe of a fly carrying GC16-GAL4 and UAS-GCaMP. The arrow indicates DC4. Peak responses of fluorescence intensity (ΔF) are shown here. Scale bars, 10 μm.

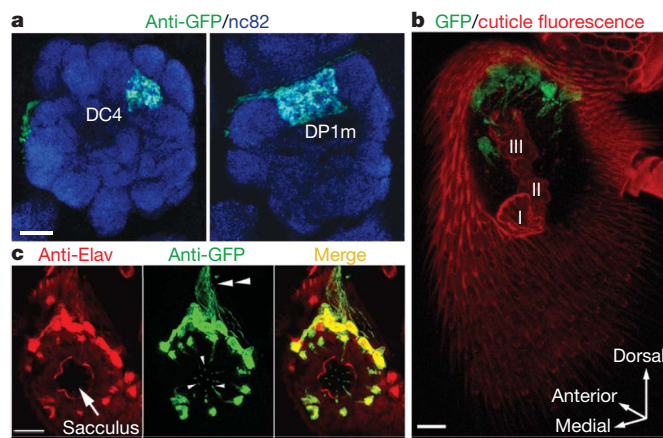
<sup>1</sup>Molecular Neurobiology Program, Skirball Institute of Biomolecular Medicine, Department of Cell Biology, New York University, School of Medicine, New York, New York 10016, USA. <sup>2</sup>Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, CH-1015 Lausanne, Switzerland. <sup>†</sup>Present addresses: Centre des Sciences du Goût et de l'Alimentation, UMR-6265 CNRS, UMR-1324 INRA, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France (Y.G.); Université Paris Diderot – Paris 7, 16 rue Marguerite Duras, 75013 Paris, France (C.L.).

\*These authors contributed equally to this work.

Because CO<sub>2</sub>, when dissolved in the lymph fluid inside the antennal sensilla that harbour OSNs, can generate metabolites, such as carbonic acid and bicarbonate ions, we tested whether DC4 could be activated by CO<sub>2</sub> metabolites. As shown in Fig. 1c, DC4 was stimulated by carbonic acid but not by bicarbonate, suggesting that these neurons detect acidosis produced by increased CO<sub>2</sub> concentrations, rather than CO<sub>2</sub> itself.

Axonal projections to DC4 originate from a population of OSNs that reside in coeloconic sensilla and express neither insect odorant receptors nor gustatory receptors. Instead, we found that these neurons express a novel receptor, IR64a, a member of the chemosensory ionotropic glutamate receptor family<sup>2</sup>. The *IR64a* promoter, *IR64a*-GAL4, driving UAS-CD8GFP, labelled the DC4 glomerulus and another glomerulus, DP1m (Fig. 2a). Anti-IR64a immunohistochemistry demonstrated that the *IR64a*-GAL4 driver recapitulated the endogenous *IR64a* expression (Supplementary Fig. 2a). We detected about 16 ± 0.9 *IR64a*+ cells (Supplementary Fig. 2b) surrounding the third chamber of the sacculus<sup>15</sup>, which is a three-chamber pit organ that opens to the posterior surface of the antenna (Fig. 2b). These *IR64a*+ cells send their dendrites to grooved sensilla that project to the interior of the sacculus (Fig. 2b, c).

Because *IR64a*+ neurons project to the DC4 and DP1m glomeruli, we determined whether only DC4, or both DC4 and DP1m, were activated by acids by calcium imaging on flies carrying *IR64a*-GAL4 and UAS-GCaMP. All acids examined, but not non-acidic odorants, activated DC4 (Fig. 3a, b and Supplementary Table 1). In contrast, DP1m was activated by acidic and non-acidic odorants (Fig. 3b and Supplementary Fig. 3). We wondered whether DP1m and DC4 might be activated by the functional side chains of some organic acids, rather than by the protons. We therefore tested whether inorganic acids such as hydrochloric acid (HCl) and nitric acid (HNO<sub>3</sub>), which dissociate completely in water and generate protons without an organic moiety, could activate DP1m and DC4. These inorganic acids, probably free protons in water vapour, activated DC4 in a dosage-dependent manner but did not activate DP1m (Fig. 3a, b). This is consistent with the observation that only DC4 is activated by CO<sub>2</sub>, which contains no associated side chains. Furthermore, the strength of the DC4 activation was inversely correlated with the pH of one odorant, sodium acetate (Fig. 3c). These results demonstrate that the neurons projecting to the DC4 glomerulus are highly specific for the detection of acidity.



**Figure 2 | DC4 is innervated by coeloconic sensillar neurons expressing *IR64a*.** **a**, Two different optical planes of the antennal lobe (DC4 (left) and DP1m (right)) of a fly bearing *IR64a*-GAL4;UAS-CD8GFP (**a**, **b**, **c**). **b**, *IR64a*+ cells in green extend their dendrites to the third chamber of the sacculus. Red autofluorescence fortuitously depicts the outline of the antenna and sacculus. I, II and III represent the first, second and third chambers, respectively. **c**, An optical plane of the antenna across the third chamber (arrow) immunostained with anti-Elav (red) and anti-GFP (green). Arrowheads, dendritic terminals; double arrowhead, axonal bundles. Scale bars, 10 μm.

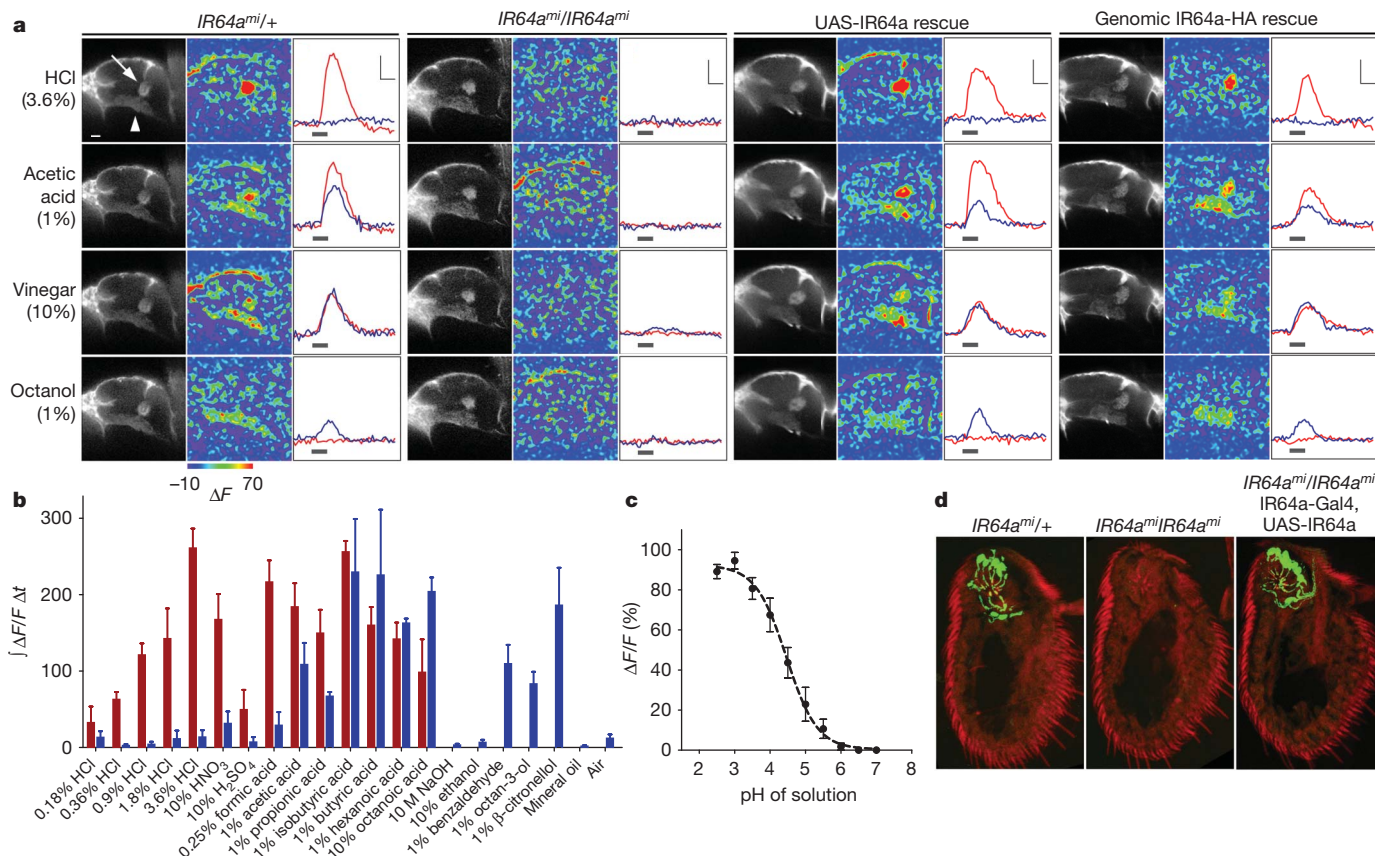
To determine whether the *IR64a* gene is required for acid detection, we obtained a mutation (*IR64a<sup>mi</sup>*) with a transposable Minos element<sup>16</sup> inserted into the third intron of the *IR64a* locus. Flies homozygous for the *IR64a<sup>mi</sup>* allele had significantly decreased *IR64a* messenger RNA transcript (Supplementary Fig. 4) and *IR64a* protein (Fig. 3d) in the antennae compared with wild-type flies. *IR64a<sup>mi</sup>* is therefore a strong loss-of-function mutation. This mutation abrogated glomerular activation of DC4 by acids (Fig. 3a and Supplementary Table 2). *IR64a<sup>mi</sup>* also attenuated the activation of DP1m to acidic and non-acidic odorants. Two different *IR64a* transgenes—the genomic *IR64a*-HA (where HA stands for haemagglutinin) driven by its own regulatory elements and a UAS-*IR64a* complementary DNA driven by *IR64a*-GAL4—rescued the odour-sensing defects of *IR64a<sup>mi</sup>* mutants (Fig. 3a and Supplementary Table 2). These results demonstrated that *IR64a* has a cell autonomous function as a component of the acid-sensing machinery required for DC4 activation and the machinery through which other odorants activate DP1m.

*IR64a* protein is localized in the cell bodies and dendrites but not in axonal processes in the antennal lobe (Figs 3d and 4e). It is highly enriched in the tip of the dendritic terminals that innervate coeloconic sensilla protruding into the lumen of the third chamber of the sacculus. The subcellular localization of *IR64a* is consistent with its direct involvement in acid detection. To determine the role of *IR64a* as a putative acid receptor, we ectopically expressed *IR64a* in another population of sensory neurons that are normally insensitive to acids and asked whether *IR64a* is capable of conferring acid sensitivity in these neurons. We expressed *IR64a* by using the *IR76a*-GAL4 driver<sup>2</sup>, which is expressed in coeloconic sensory neurons that project to the VM4 glomerulus in the antennal lobe<sup>2</sup>. Calcium imaging experiments showed that ectopic expression of *IR64a* induces odour sensitivity in VM4 to organic acids and octan-3-ol, which normally activate DP1m (Supplementary Fig. 5), substantiating the notion that *IR64a* is the direct determinant in odour detection. However, *IR64a* alone was not capable of conferring sensitivity to DC4-specific stimuli such as inorganic acids or CO<sub>2</sub> (Supplementary Fig. 5 and Supplementary Table 2). This result suggests that although *IR64a* alone can induce responsiveness to several odorants that activate DP1m, it probably requires a co-receptor in DC4 neurons to mediate the specificity to acidity.

Having shown that *IR64a* is part of the acid-sensing machinery, we next determined whether *IR64a*+ neurons are necessary for the flies' behavioural response to acids. We engineered flies in which *IR64a*+ cells were silenced by the targeted expression of tetanus toxin (TNT)<sup>17</sup>. In a T-maze, these flies showed significant decreases in avoidance to several acids, whereas responses to non-acidic odorants such as benzaldehyde and octan-3-ol were unaffected (Fig. 4a). However, this experiment could not determine whether it is DC4 or DP1m that is required for acid avoidance, because *IR64a*-GAL4 is expressed in both populations of sensory neurons. Nonetheless, DP1m is unlikely to be important because it is not activated by acidity (Fig. 3a–c). To confirm the importance of DC4 in acid avoidance, we generated a GAL80 transgene under the control of the *IR64a* promoter and crossed this line to flies carrying GC16-GAL4 and UAS-*Shibire<sup>ts</sup>*. Because GC16-GAL4 is expressed in DC4 neurons, but not in DP1m neurons, the *IR64a*-GAL80 transgene selectively relieves neuronal inhibition only in DC4 neurons. We confirmed that *IR64a*-GAL80 suppresses GC16-GAL4 activity only in DC4 neurons by using a UAS-CD8GFP transgene (Supplementary Fig. 6a). The *IR64a*-GAL80 transgene rescued the behavioural defects of flies carrying GC16-GAL4 and UAS-*Shibire<sup>ts</sup>*, supporting the specific role of DC4 in mediating behavioural responses to acids (Supplementary Fig. 6b).

We next examined whether artificial activation of *IR64a*+ neurons is sufficient to trigger avoidance responses. We generated *GR63a<sup>1</sup>;IR64a<sup>mi</sup>* double mutants expressing the CO<sub>2</sub> receptors UAS-GR21a and UAS-GR63a, and a calcium-sensitive GFP, UAS-GCaMP, by using the *IR64a*-GAL4 driver. Expression of the two CO<sub>2</sub> receptors in CO<sub>2</sub>-insensitive





**Figure 3 | Activation of DC4 by acidity requires *IR64a*.** **a**, *In vivo* calcium imaging of *IR64a-GAL4;UAS-GCaMP* flies with different genotypes. For each genotype, pre-stimulation (left), peak  $\Delta F$  responses (middle) and traces for glomerular activation (right) are shown. Arrow, DC4; arrowhead, DP1m. In the traces, the vertical scale indicates 25%  $\Delta F/F$  and the horizontal scale 1 s. The horizontal black bar below each trace indicates the duration of odorant exposure. Red traces, DC4; blue traces, DP1m. **b**, Integration of the GCaMP

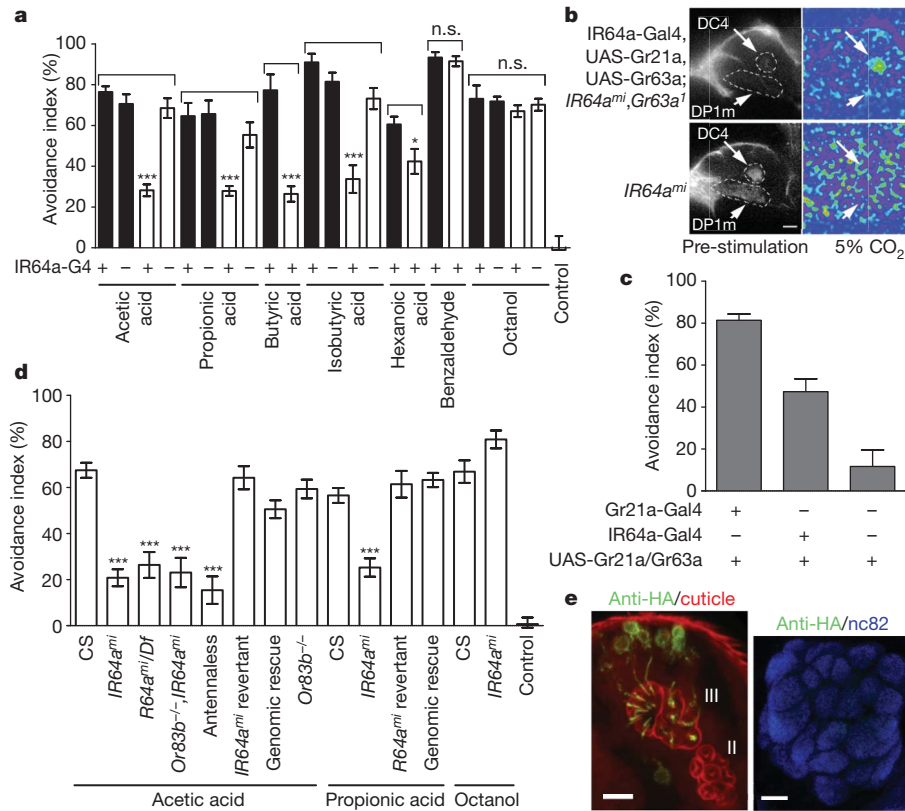
signals over time during glomerular activation (see Methods). Error bars indicate s.e.m. ( $n = 4-12$ ). Red, DC4; blue, DP1m. **c**, The GCaMP signal of DC4 responding to sodium acetate solutions of different pH values. The dotted line is a nonlinear regression fit of the data:  $R^2 = 0.902$ . Error bars indicate s.e.m. ( $n = 6$ ). **d**, Anti-*IR64a* (green) immunohistochemistry on sectioned antennae. Red auto-fluorescence outlines the antenna. Scale bars, 10  $\mu$ m.

sensory neurons was previously shown to be sufficient to confer ectopic sensitivity to  $\text{CO}_2$ <sup>8,9</sup>. Indeed, the DC4 glomerulus in these flies was artificially stimulated by 5%  $\text{CO}_2$  (Fig. 4b). However, we could not detect the activation of DP1m by  $\text{CO}_2$  in these flies (Fig. 4b), possibly because *GR21a* and *GR63a* receptors do not function properly in DP1m neurons. Behavioural experiments demonstrated that *GR63a<sup>1</sup>;IR64a<sup>mi</sup>* double mutant flies failed to distinguish ambient air from 5%  $\text{CO}_2$ . However, the  $\text{CO}_2$ -blind flies with  $\text{CO}_2$  receptors expressed in *IR64a+* neurons showed robust avoidance to  $\text{CO}_2$  (Fig. 4c). This suggests that avoidance behaviour is hardwired into the olfactory circuitry that detects acidity. Because DP1m in these flies does not seem to be activated by  $\text{CO}_2$ , we reason that activation of DC4 neurons alone is sufficient for generating avoidance responses. These data, together with the observation that acidity evoked calcium responses only in DC4, firmly establish that *IR64a+* neurons projecting to the DC4 glomerulus are necessary for acid sensation and sufficient for avoidance behaviour. These results provide strong evidence for the functional segregation of acid sensing at the periphery that drives innate avoidance behaviour.

Consistent with the physiological defects was the observation that *IR64a<sup>mi</sup>* flies had impaired avoidance to acids but normal responses to an unrelated odorant (Fig. 4d). Conversely, flies in which the Minos element was precisely excised from the *IR64a* locus (*IR64a* revertants) and those carrying an *IR64a* transgene in the *IR64a<sup>mi</sup>* mutant background showed robust avoidance to acids (Fig. 4d). Although the average avoidance indices of *IR64a<sup>mi</sup>* and *IR64a-GAL4*  $\times$  *UAS-TNT* flies were significantly different from those of the wild type, they still showed moderate avoidance responses to acids (avoidance index of

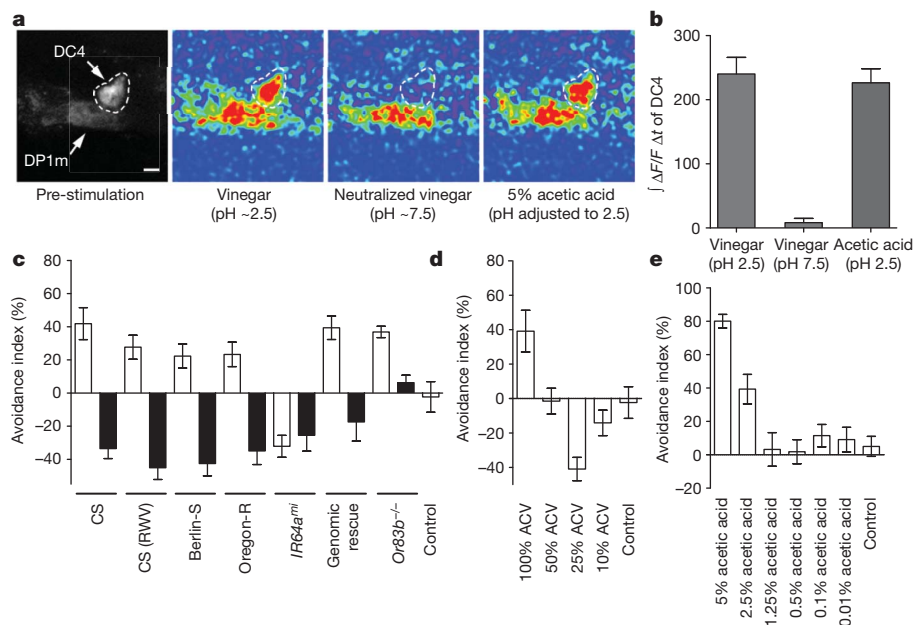
20–25%). This residual response is unlikely to be mediated by the olfactory system, because flies lacking antennae had avoidance responses similar to those of *IR64a<sup>mi</sup>* (Fig. 4d). Thus, additional acid sensors probably exist elsewhere in the fly.

Fruit flies are often called ‘vinegar flies’ because of their attraction to vinegar. Indeed, flies were attracted to certain concentrations of vinegar in a T-maze (Fig. 5d). However, a major ingredient of vinegar is acetic acid, which flies avoid (Fig. 5e). It is possible that flies are not repelled by vinegar because other constituents in vinegar inhibit DC4 activation by acetic acid. Alternatively, constituents other than acetic acid in vinegar might elicit an attraction response that overrides DC4-mediated avoidance. To distinguish between these possibilities, we performed *in vivo* calcium imaging to measure the activation of DC4 after exposure to vinegar. As shown in Fig. 5a, b, apple cider vinegar (ACV), which contains about 5% acetic acid, activated DC4 as effectively as pure 5% acetic acid. These results suggest that vinegar contains attractants capable of overcoming DC4-mediated avoidance by activating other olfactory receptors. We predicted that neutralized vinegar would not activate DC4 and should be more attractive to flies because it still contains attractants. Consistent with this prediction, calcium imaging showed that DC4 was not stimulated by neutralized vinegar (Fig. 5a, b). Moreover, wild-type flies avoided a high concentration of vinegar (pH 2.5) in a T-maze but became attracted to neutralized vinegar (pH 7.5) at the same concentration (Fig. 5c). A similar behavioural switch was observed in other *D. melanogaster* strains such as Berlin and Oregon R, and with another type of vinegar (Fig. 5c). Furthermore, flies were attracted to 25% ACV (Fig. 5d), but not to acetic acid that had



**Figure 4 | IR64a<sup>+</sup> neurons and *IR64a* are necessary and sufficient for avoidance behaviour.** **a**, Avoidance responses of flies expressing TNT (open bars) or inactivated TNT (filled bars) by *IR64a-GAL4* in a T-maze. Error bars indicate s.e.m. (n = 13–37). **b**, Calcium imaging of *GR63a<sup>1</sup>;IR64a<sup>mi</sup>* mutant carrying *IR64a-GAL4*, *UAS-GR21a*, *UAS-GR63a* and *UAS-GCaMP* in response to 5% CO<sub>2</sub>. **c**, Avoidance of *GR63a<sup>1</sup>;IR64a<sup>mi</sup>* mutant expressing CO<sub>2</sub>

receptors to 5% CO<sub>2</sub>. **d**, Avoidance of flies with different genotypes blindly tested in a T-maze. Df, genomic deficiency uncovering *IR64a* locus. CS, Canton S. Error bars indicate s.e.m. (n = 12–27). **e**, Anti-HA immunohistochemistry on a fly harbouring an *IR64a-HA* genomic transgene. *IR64a-HA* protein (green) is localized in dendrites (left), but not in axons (right). Scale bars, 10 μm. Asterisk, P < 0.05; three asterisks, P < 0.001 (analysis of variance, Tukey test).



**Figure 5 | Fruit flies are attracted to components other than acetic acid in vinegar.** **a**, Calcium imaging of flies bearing *IR64a-GAL4* and *UAS-GCaMP*. **b**, Integration of the GCaMP signals during odour presentation. Error bars indicate s.e.m. (n = 9). **c**, Responses of starved flies to regular (pH about 2.5; open bars) and neutralized (pH about 7.5; filled bars) vinegar in a T-maze. RWV, red wine vinegar. Error bars indicate s.e.m. (n = 8–22). Positive

avoidance index indicates avoidance; negative index shows attraction. **d**, Starved flies are attracted to diluted ACV (25%). Error bars indicate s.e.m. (n = 6–11). **e**, Starved flies are not attracted to diluted acetic acid. The acidity of 25% ACV is the same as that of 1.25% acetic acid. Error bars indicate s.e.m. (n = 5–8).

been diluted to the same concentration of acidity (Fig. 5e). This further supports the model that flies are attracted to components other than acid in vinegar. Avoidance of vinegar requires a functional IR64a+ circuit, because *IR64a<sup>mi</sup>* mutants were equally attracted to vinegar and to neutralized vinegar (Fig. 5c).

Animals across various phyla show innate aversion to a plume of acid<sup>18–20</sup>, often emanating from spoiled food or unripe fruit. Our characterization of *Drosophila* IR64a provides a cellular and molecular mechanism that can explain the distinct olfactory sensation of acidity. In the mammalian taste system, acid detection is mediated by a unique cell type, independently of other taste modalities<sup>21</sup>. This labelled-line organization is similar to those of acid and CO<sub>2</sub> receptors in the fly olfactory system. Both acid and CO<sub>2</sub> sensors are highly specific to their ligands and mediate similar avoidance behaviour. This raises a further question: where are these two similar aversive stimuli represented in the brain? The identification of neural substrates in the central nervous system mediating acid and CO<sub>2</sub> sensation will facilitate future mapping of the avoidance circuitry.

## METHODS SUMMARY

**Transgenic flies and fly stocks.** IR64a-GAL4 was made by cloning the 5' sequence of *IR64a* into pCasper4-AUG-GAL4X (ref. 22). The IR64a-HA genomic transgene was constructed in pCasper4 with an in-frame HA-coding sequence at the carboxy terminus of *IR64a*. UAS-IR64a was made by cloning *IR64a* cDNA into pUAST vector. *IR64a<sup>mi</sup>* flies and flies carrying a genomic deficiency uncovering the *IR64a* locus were obtained from the Bloomington *Drosophila* Stock Center.

**Antibody and immunohistochemistry.** Rabbit anti-IR64a polyclonal antibody was generated against a peptide antigen (SGKRDDGEMEEEEPPGQQ). Immunostaining of the fly brain and cryosectioning of the antennae were performed as described previously<sup>7,23</sup>.

**Calcium imaging.** *In vivo* calcium imaging was conducted with a live, behaving fly preparation that was subjected to a minimally invasive surgical procedure<sup>12</sup>. In brief, flies were glued to a custom-made plastic slide. Head cuticle was removed to expose the dorsal side of the brain, which was submerged in adult-haemolymph-like buffer<sup>4</sup>. The glomerular responses of the antennal lobe were detected by two-photon microscopy. Odour from the headspace of odorant-containing vials was delivered to the fly antenna by a puffing device.

**Behavioural test.** Testing of acid-evoked behavioural responses of flies was conducted in a T-maze as described previously<sup>7</sup>. For the experiments with vinegars, flies were starved for 23 h before behaviour testing.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 12 January; accepted 27 September 2010.**

**Published online 17 November 2010.**

1. Dravnieks, A. Odor quality: semantically generated multidimensional profiles are stable. *Science* **218**, 799–801 (1982).
2. Benton, R., Vannice, K. S., Gomez-Diaz, C. & Vosshall, L. B. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* **136**, 149–162 (2009).
3. Ng, M. *et al.* Transmission of olfactory information between three populations of neurons in the antennal lobe of the fly. *Neuron* **36**, 463–474 (2002).
4. Wang, J. W., Wong, A. M., Flores, J., Vosshall, L. B. & Axel, R. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112**, 271–282 (2003).
5. Kellogg, F. E. Water vapour and carbon dioxide receptors in *Aedes aegypti*. *J. Insect Physiol.* **16**, 99–108 (1970).
6. Thom, C., Guerenstein, P. G., Mechaber, W. L. & Hildebrand, J. G. Floral CO<sub>2</sub> reveals flower profitability to moths. *J. Chem. Ecol.* **30**, 1285–1288 (2004).
7. Suh, G. S. *et al.* A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* **431**, 854–859 (2004).

8. Jones, W. D., Cayirlioglu, P., Kadow, I. G. & Vosshall, L. B. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**, 86–90 (2007).
9. Kwon, J. Y., Dahanukar, A., Weiss, L. A. & Carlson, J. R. The molecular basis of CO<sub>2</sub> reception in *Drosophila*. *Proc. Natl Acad. Sci. USA* **104**, 3574–3578 (2007).
10. Faucher, C., Forstreuter, M., Hilker, M. & de Bruyne, M. Behavioral responses of *Drosophila* to biogenic levels of carbon dioxide depend on life-stage, sex and olfactory context. *J. Exp. Biol.* **209**, 2739–2748 (2006).
11. Kitamoto, T. Conditional modification of behavior in *Drosophila* by targeted expression of a temperature-sensitive shibire allele in defined neurons. *J. Neurobiol.* **47**, 81–92 (2001).
12. Yoro, S. *et al.* Distinct sensory representations of wind and near-field sound in the *Drosophila* brain. *Nature* **458**, 201–205 (2009).
13. Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca<sup>2+</sup> probe composed of a single green fluorescent protein. *Nature Biotechnol.* **19**, 137–141 (2001).
14. Couto, A., Alenius, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr. Biol.* **15**, 1535–1547 (2005).
15. Shambhag, S. R., Singh, K. & Singh, R. N. Fine structure and primary sensory projections of sensilla located in the sacculus of the antenna of *Drosophila melanogaster*. *Cell Tissue Res.* **282**, 237–249 (1995).
16. Franz, G., Loukeris, T. G., Dialektaki, G., Thompson, C. R. & Savakis, C. Mobile Minos elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proc. Natl Acad. Sci. USA* **91**, 4746–4750 (1994).
17. Sweeney, S. T., Broadie, K., Keane, J., Niemann, H. & O'Kane, C. J. Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects. *Neuron* **14**, 341–351 (1995).
18. Sambongi, Y. *et al.* *Caenorhabditis elegans* senses protons through amphid chemosensory neurons: proton signals elicit avoidance behavior. *Neuroreport* **11**, 2229–2232 (2000).
19. Joseph, R. M., Devineni, A. V., King, I. F. & Heberlein, U. Oviposition preference for and positional avoidance of acetic acid provide a model for competing behavioral drives in *Drosophila*. *Proc. Natl Acad. Sci. USA* **106**, 11352–11357 (2009).
20. Ziemann, A. E. *et al.* The amygdala is a chemosensor that detects carbon dioxide and acidosis to elicit fear behavior. *Cell* **139**, 1012–1021 (2009).
21. Huang, A. L. *et al.* The cells and logic for mammalian sour taste detection. *Nature* **442**, 934–938 (2006).
22. Vosshall, L. B., Wong, A. M. & Axel, R. An olfactory sensory map in the fly brain. *Cell* **102**, 147–159 (2000).
23. Larsson, M. C. *et al.* Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**, 703–714 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Dus for helpful discussions and technical support for making anti-IR64a antibody; M. Kim for technical assistance; J. Treisman, N. Ringstad, C. Desplan, M. Warman, G. Fishell, M. Chesler, A. Wong, L. Vosshall and D. Anderson for discussions and comments on the manuscript; and R. Lehmann for sharing her two-photon microscope. This work was initiated in D. Anderson's laboratory at California Institute of Technology. G.S.B.S. thanks H.-R. Song for support. Financial support was provided by a National Research Service Award fellowship (M.A.), the Boehringer Ingelheim Foundation (R. Bell), the Centre National de la Recherche Scientifique (Y.G.), a European Research Council Starting Independent Researcher Grant (R. Benton), the Swiss National Science Foundation (R. Benton), the Alfred P. Sloan Foundation (G.S.), the Whitehall Foundation (G.S.), a Whitehead President Award (G.S.) and NIH grant 1R01GM089746 (G.S.).

**Author Contributions** G.S.B.S. identified the GC16-GAL4 line from a screen and performed initial imaging experiments. M.A. generated most of the transgenic flies, performed calcium imaging and some behavioural experiments, and conducted immunohistochemistry. S.M. performed most of the behavioural testing and immunohistochemistry. C.L. assisted in characterizing *IR64a* mutant alleles. Y.G., R. Bell and R. Benton generated IR64a-GAL4 and UAS-IR64a transgenic flies and showed that IR64a+ neurons innervate the DC4 and DP1m glomeruli. M.A. and G.S.B.S. analysed and interpreted the data and wrote the manuscript with inputs from all the authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.S.B.S. ([greg.suh@med.nyu.edu](mailto:greg.suh@med.nyu.edu)).



## METHODS

**Fly stocks.** Flies were raised on standard cornmeal medium at 25 °C. *IR64a<sup>mi</sup>* (stock number: 24610) and deficiency flies (stock number: 25119) uncovering the *IR64a* locus were obtained from the Bloomington stock centre. *IR64a* revertant flies were generated by mobilizing the Minos element with the use of Minos transposase as described previously<sup>24</sup>. Precise excision lines were identified by a loss of GFP signal in the eyes and confirmed by PCR genotyping and sequencing. Other flies were described previously: UAS-GCaMP(1.3) (ref. 4), *Or83b<sup>-/-</sup>* (ref. 23), *Or83b-GAL4* (ref. 4), *Gr63a<sup>1</sup>* (ref. 8), UAS-*Shi<sup>ts</sup>* (ref. 11), UAS-TNT and UAS-ImpTNT (ref. 17).

**Transgenic constructs and flies.** *IR64a-GAL4* was made by cloning a 2.5-kilobase (kb) or 8-kb DNA sequence upstream of *IR64a* into pCasper4-AUG-GAL4X (ref. 22). The 8-kb *IR64a* upstream sequence was subcloned into pCasper-GAL80 vector to generate *IR64a-GAL80*. The *IR64a-HA* genomic duplication construct was made in pCasper4 with an 8-kb *IR64a* upstream sequence, a 4-kb *IR64a* genomic coding sequence (including introns), and an in-frame HA coding sequence followed by a 1.4-kb *IR64a* downstream genomic sequence. UAS-*IR64a* was made by cloning *IR64a* cDNA sequence into pUAST vector. Transgenic flies were generated by Bestgene, Inc.

**Antibodies and immunohistochemistry.** Rabbit anti-*IR64a* polyclonal antibody was generated against a peptide antigen (SGKRDDGEMEEEEPPGQQ) corresponding to 202–220 amino-acid residues of *IR64a* protein by Yenzyme Inc. This peptide is located within the predicted amino-terminal extracellular domain. Anti-*IR64a* serum was affinity purified and used at 1:1,000 dilutions for immunohistochemistry. Other antibodies used in immunohistochemistry were monoclonal anti-HA (1:1,000 dilution; Covance), rabbit anti-GFP (1:500 dilution; Invitrogen), monoclonal anti-Elav (1:100 dilution; Hybridoma Bank) and nc82 (1:50 dilution; Hybridoma Bank). Immunostaining of the fly brain and cryosectioning of the antennae were performed as described previously<sup>7,23</sup>.

**Preparation and delivery of odorants.** Odorants and acids were dissolved in water or mineral oil (percentages shown are v/v). Odorants and acids that were diluted in water include hydrochloric acid, nitric acid, sulphuric acid, vinegar, formic acid, acetic acid, propionic acid, butyric acid, isobutyric acid, hexanoic acid, sodium hydroxide, methanol, ethanol, formaldehyde, paraformaldehyde, acetaldehyde, propionaldehyde, butyraldehyde and valeraldehyde. Carbonic acid was freshly made from sodium bicarbonate solution (1 M) and adjusted to pH 2.0 by the addition of hydrochloric acid. Sodium acetate solution (100 mM) was made in water and was adjusted to the desired pH by the addition of acetic acid. Other odorants were dissolved in mineral oil. Diluted odorants (300 µl) or control solvent were placed in 16-ml glass vials with a sealed rubber cap (Supelco) and allowed to equilibrate with the head space for at least 1 h before use.

The air within the head space of the odorant-containing vials was delivered to the fly antennae by using a puffing device, which includes a pump generating a 22 ml min<sup>-1</sup> flow of humidified air and an electronic valve controller. The puffing device was programmed for precise control of the on and off switch between valves connected to the odorant-containing vials. In the resting state, air constantly flows through the control vial and is delivered to the fly antennae. On stimulation with odour, the valve connected to an odorant-containing vial opens for 1 s to allow odorant in the head space of the vial to be delivered to the fly antennae. After odorant delivery, the valve quickly shuts down and redirects the air to flow through the control vial again. This valve switching delivery system allows fast clearance of odorants and produces minimal mechanical disturbance.

**Live fly preparation and calcium imaging.** Flies were anaesthetized and glued to a custom-made plastic slide by their wings, with the use of ultraviolet-convertible glue (Kemxert Corp). The proboscis was glued to the chest to restrain head movement. The head and thorax were pushed through an opening (0.8 mm wide and 1.6 mm long) and exposed to the upper side of the slide. The fly was carefully oriented such that the antennae pointed downward. Small drops of wax (about 55 °C) were applied around the eyes and thorax to immobilize the fly. Head cuticle was carefully removed to expose the dorsal side of the brain, which was submerged in adult-haemolymph-like buffer<sup>4</sup>. This protocol is more sensitive than the previously described dissected brain preparation, which was used to show that CO<sub>2</sub> solely activated the V glomerulus<sup>7</sup>.

Glomerular responses to odour stimulations were recorded by using a two-photon microscope with a 40× water-immersion objective lens. Real-time images were acquired at 7.57 frames per second with a resolution of 128 pixels × 128 pixels. Imaging data were processed with the use of ImageJ with a custom plug-in to generate pseudo-colour intensity images.  $\Delta F/F$  was calculated as described in ref. 25.  $|\Delta F/F| \Delta t$  was computed as the total area under activation peak divided by the width of the peak.

Flies used in the imaging experiments were 10–15 days old with two exceptions: first, 5-day-old UAS-GCaMP(1.3);GC16-GAL4 flies were used in Fig. 1c because GC16-GAL4 expression in the DC4 glomerulus becomes very weak in older flies; second, the presence of many UAS elements significantly decreased the efficiency of *IR64a-GAL4* to drive UAS-GCaMP expression. Thus, older (30-day-old) flies carrying UAS-GCaMP;*IR64a-GAL4*/UAS-*Gr21a*;*Gr63a<sup>1</sup>*, *IR64a<sup>mi</sup>*, UAS-*Gr63a* were used in Fig. 4b. Flies described in Fig. 3a were *IR64a<sup>mi</sup>*/+ (UAS-GCaMP;*IR64a-GAL4*;*IR64a<sup>mi</sup>*, *IR64a-GAL4*/TM6B), *IR64a<sup>mi</sup>*/*IR64a<sup>mi</sup>* (UAS-GCaMP;*IR64a-GAL4*;*IR64a<sup>mi</sup>*, *IR64a-GAL4*), UAS-*IR64a* rescue (UAS-GCaMP;UAS-*IR64a*/*IR64a-GAL4*;*IR64a<sup>mi</sup>*, *IR64a-GAL4*) and genomic *IR64a-HA* rescue (UAS-GCaMP;*IR64a-GAL4*;*IR64a<sup>mi</sup>*, *IR64a-HA*).

**Behavioural tests.** Flies 7–12 days old were used for most of the behavioural tests, except that 5-day-old GC16-GAL4 flies were used in Fig. 1a. For antennaless flies, behavioural tests were performed 24 h after surgery. All behavioural tests were performed at 23–25 °C with the exception of experiments with UAS-*Shi<sup>ts</sup>* flies: flies carrying UAS-*Shi<sup>ts</sup>* were heat-shocked in a 34 °C water bath for 4–5 min, and subsequent behavioural testing was done at 29 °C. T-maze experiments were performed as described previously<sup>7</sup>. Odorants for T-maze tests were diluted in water or mineral oil to the following concentrations (v/v): acetic acid (10%), propionic acid (2.5%), isobutyric acid (1.25%), butyric acid (5%), hexanoic acid (5%), benzaldehyde (1.25%) and octan-3-ol (5%). ACV and red wine vinegar were used without dilution. Neutralized vinegars were obtained by the addition of sodium hydroxide to vinegars until the pH reached 7.5. Flies previously starved for 23 h were used for behavioural testing with vinegar.

The avoidance index was calculated as (the number of flies in control tube minus the number of flies in experimental tube) divided by (the number of flies in the experimental tube plus the number of flies in the control tube). Flies being tested were given a choice in a T-maze for 30–40 s before the flies in each tube were counted. A group of 30–35 flies was tested in each trial.

24. Metaxakis, A., Oehler, S., Klinakis, A. & Savakis, C. Minos as a genetic and genomic tool in *Drosophila melanogaster*. *Genetics* **171**, 571–581 (2005).

25. Asahina, K., Louis, M., Piccinotti, S. & Vosshall, L. B. A circuit supporting concentration-invariant odor perception in *Drosophila*. *J. Biol.* **8**, 9 (2009).

# Interdependence of behavioural variability and response to small stimuli in bacteria

Heungwon Park<sup>1</sup>, William Pontius<sup>2</sup>, Calin C. Guet<sup>3</sup>, John F. Marko<sup>4</sup>, Thierry Emonet<sup>2</sup> & Philippe Cluzel<sup>3</sup>

**The chemotaxis signalling network in *Escherichia coli* that controls the locomotion of bacteria is a classic model system for signal transduction<sup>1,2</sup>. This pathway modulates the behaviour of flagellar motors to propel bacteria towards sources of chemical attractants. Although this system relaxes to a steady state in response to environmental changes, the signalling events within the chemotaxis network are noisy and cause large temporal variations of the motor behaviour even in the absence of stimulus<sup>3</sup>. That the same signalling network governs both behavioural variability and cellular response raises the question of whether these two traits are independent. Here, we experimentally establish a fluctuation–response relationship in the chemotaxis system of living bacteria. Using this relationship, we demonstrate the possibility of inferring the cellular response from the behavioural variability measured before stimulus. In monitoring the pre- and post-stimulus switching behaviour of individual bacterial motors, we found that variability scales linearly with the response time for different functioning states of the cell. This study highlights that the fundamental relationship between fluctuation and response is not constrained to physical systems at thermodynamic equilibrium<sup>4</sup> but is extensible to living cells<sup>5</sup>. Such a relationship not only implies that behavioural variability and cellular response can be coupled traits, but it also provides a general framework within which we can examine how the selection of a network design shapes this interdependence.**

It is standard procedure to characterize the stochastic dynamics of physical systems in thermodynamic equilibrium by measuring spontaneous fluctuations and responses to small external perturbations. Because these two distinct measurements contain the same information, they are related by the fluctuation-dissipation theorem<sup>4</sup>. Although the fluctuation-dissipation theorem has practical applications—to evaluate force-extension sensors for single biomolecules<sup>6,7</sup> and to predict static cell-to-cell variability of gene expression<sup>8,9</sup>—it has not been possible to apply it directly to the study of the dynamical behaviour of living cells because they are open systems with significant non-thermal dynamics. However, this theorem has recently been extended to a fluctuation-response theorem (FRT) for systems that are not in thermodynamic equilibrium but that have a well-defined steady state and Markovian dynamics<sup>5,10–12</sup>. For application to living cells this condition amounts to studying dynamic processes with sufficiently short ‘memory’ that they can relax to a well-defined steady state. Here we use the FRT as an operational framework to establish the interdependence of distinct cellular traits, such as cellular fluctuations and response to a small stimulus, without relying on the biochemical details of a specific signalling pathway. To tackle this question experimentally, we used the well-characterized chemotaxis system in *E. coli*, which governs bacterial locomotion<sup>13</sup>.

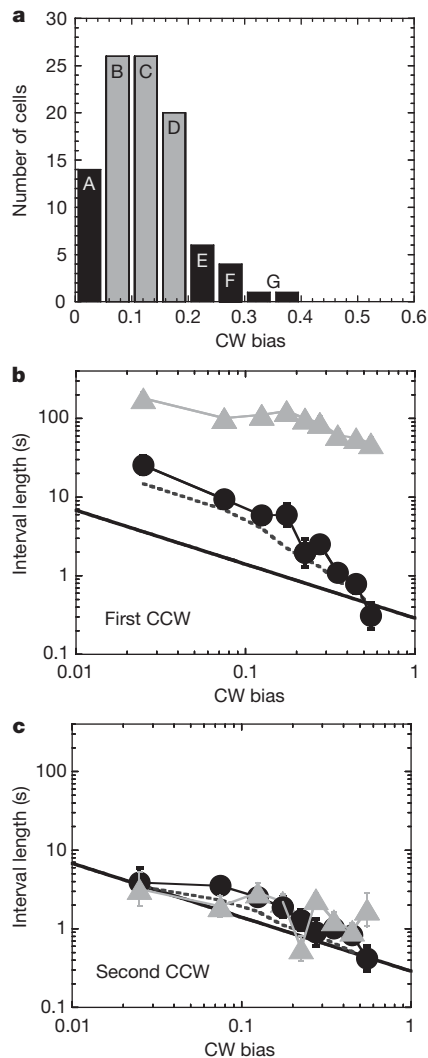
The chemotaxis network regulates the rotation direction—clockwise (CW) or counter-clockwise (CCW)—of the flagellar motors, which control the swimming direction of the cell<sup>1,2</sup>. One of the hallmarks of

bacterial chemotaxis is adaptation. Following a stepwise stimulus, the CW bias (the probability that the motor will rotate clockwise) decreases abruptly, before slowly adapting back to its pre-stimulus level. Even when bacteria are adapted to their environment, the CW bias of individual cells fluctuates around the mean. These temporal fluctuations in CW bias reflect slow fluctuations in signalling events throughout the transduction network<sup>14</sup>. To verify that the bacterial chemotaxis system satisfies the FRT, we monitored both the temporal fluctuations of the CW bias before stimulus and the cellular response to a small stimulus at the single-cell level. Both quantities were obtained from the time series of CW and CCW intervals of individual motors from bacteria immobilized on a glass coverslip<sup>15</sup> and submerged in a motility medium that does not support growth. Such single-cell experiments are complicated by inherent cell-to-cell differences in relative chemotaxis protein concentration, leading to differences in switching dynamics (Fig. 1a). To compare cells with similar behaviour, we sorted wild-type cells according to their steady-state CW bias (Fig. 1a). These CW bias bins define different classes of cells, which, despite being genetically identical, have different dynamics and must be analysed separately<sup>3</sup>.

First, we quantified the response in single cells by measuring the length of successive CCW intervals immediately following the stimulus. The stimulus (10 nM of aspartate) used in this study is small and close to the limit of sensitivity of the bacterial chemotaxis system<sup>16</sup>. Given that CCW interval length is a stochastic variable, we averaged the CCW interval lengths after stimulus between cells and found that the mean length of the first CCW interval following stimulus was slightly longer than the mean pre-stimulus CCW interval length (Fig. 1b). Therefore, we expected the response of the system to be within the linear regime, which was necessary to apply the FRT. We also tested the response of the chemotaxis system for a stimulus 100 times larger (1  $\mu$ M aspartate). At the single-cell level, the length of the first CCW interval following the small stimulus (Supplementary Fig. 1a) was distributed around the mean CCW interval length before stimulus (Supplementary Fig. 1b). Surprisingly, the second CCW interval following the stimulus returned to near pre-stimulus length for both large and small attractant concentrations (Fig. 1c). Although the cellular response to stimulus extends in some cases beyond the second interval (Supplementary Fig. 1d, e), these results qualitatively indicate that the first CCW interval contains most of the chemotactic response to both small and large stimuli.

To characterize the system quantitatively, we defined the response time of a single cell as the cumulative length of post-stimulus CCW intervals that are strictly longer than the mean CCW interval length before stimulus (Fig. 1b, c and Supplementary Fig. 1e; see Methods for definition of response time). This procedure yields a reasonable estimate of the response time under the condition of small stimulus (Supplementary Fig. 2). We found that the response time averaged over CW bias bins decreased with CW bias for both small (Fig. 2a) and large stimuli (Fig. 2a, inset). Because all cells returned to their pre-stimulus behaviour

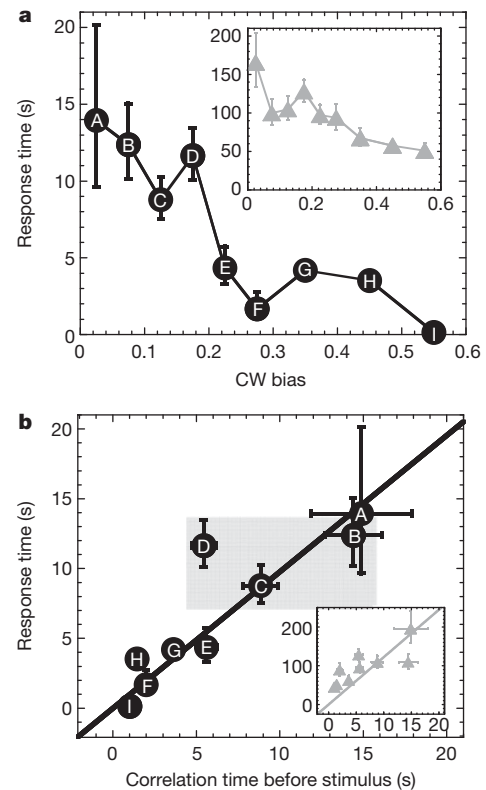
<sup>1</sup>The James Franck Institute, The Institute for Biophysical Dynamics, and The Department of Physics, University of Chicago, Chicago, Illinois 60637, USA. <sup>2</sup>Department of Molecular, Cellular, and Developmental Biology and Department of Physics, Yale University, New Haven, Connecticut 06520, USA. <sup>3</sup>FAS Center for Systems Biology, Department of Molecular and Cellular Biology, and School of Engineering and Applied Sciences, Harvard University, 52 Oxford Street, Cambridge, Massachusetts 02138, USA. <sup>4</sup>Department of Molecular Biosciences and Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA.



**Figure 1 | CCW interval lengths pre- and post-stimulus.** **a**, Histogram of CW bias of wild-type RP437 cells. We sorted cells into CW bias intervals by their pre-stimulus CW bias: 0.00–0.05 (A), 0.05–0.10 (B), 0.10–0.15 (C), 0.15–0.20 (D), 0.20–0.25 (E), 0.25–0.30 (F), 0.30–0.40 (G), 0.40–0.50 (H) and 0.50–0.60 (I). Grey bars are cells representative of wild-type behaviour. To increase the chance of obtaining cells with CW bias higher than 0.2, we transformed wild-type cells with pZE21-CheR (Methods). This extended the range of CW bias considered in our study to values greater than 0.4: bins H and I (not shown). **b**, **c**, The first (**b**) and second (**c**) mean post-stimulus CCW interval lengths versus pre-stimulus CW bias for all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR). (See Supplementary Fig. 1 for individual cells.) Black circles, cells exposed to a small stimulus (10 nM L-aspartate). Grey triangles, cells exposed to a large stimulus (1  $\mu$ M L-aspartate). Error bars show the standard error associated with the average CCW interval length in each bin. Dark grey dashed line, geometric mean of the CCW interval lengths following a randomly chosen time point in non-stimulated cells. Black line, power-law fit of the geometric mean of pre-stimulus CCW interval lengths calculated over 1,500 s for all cells (wild-type RP437 and RP437 expressing extra CheR from pZE21-CheR) as a function of the pre-stimulus CW bias (Supplementary Fig. 1b).

(Supplementary Fig. 1), the system exhibited near-precise adaptation at the single-cell level, regardless of CW bias (Supplementary Fig. 3). This result agrees with that obtained from population measurements<sup>17,18</sup> and shows that the dynamics have sufficiently short ‘memory’ and that individual cells can relax to a well-defined steady state.

A direct consequence of the linear approximation is that the response time of the system to a small external stimulus should be proportional to the correlation time of the spontaneous fluctuations before stimulus. Using serial correlation analysis<sup>19,20</sup>, we evaluated

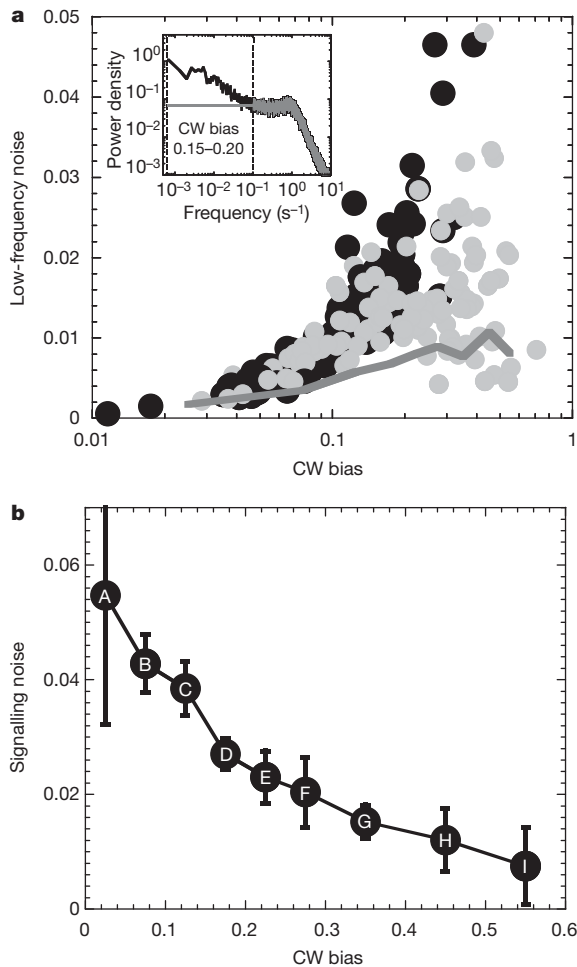


**Figure 2 | Relationship between response to stimulus and fluctuations before stimulus.** **a**, Average response time for all cells (wild-type RP437 and RP437 expressing extra CheR from pZE21-CheR) exposed to a stepwise small stimulus (black circles, 10 nM L-aspartate) or large stimulus (grey triangles in inset to **a**, 1  $\mu$ M L-aspartate). The letters correspond to the CW bias bins (Fig. 1a). Error bars show the standard error associated with the average response time within each bin. **b**, Average response time to a small stimulus (black circles) or large stimulus (grey triangles in inset to **b**) as a function of the correlation time for all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR). For the large stimulus, the average response time was adjusted by a correction factor (Supplementary Fig. 2c). The solid lines are linear fit functions forced through the origin. For the black line: response time =  $C \times$  correlation time.  $C \approx 0.98 \pm 0.10$  ( $R^2 = 0.75$ ). For the grey line in the inset: relaxation time =  $C \times$  correlation time.  $C \approx 12.23 \pm 1.83$  ( $R^2 = 0.07$ ). Error bars for the correlation time are the half-lengths of the first uncorrelated CCW intervals. Error bars for the response time are the standard error associated with the average response time within each bin. Grey area, representative behaviour of a wild-type population. Insets in **a** and **b** share axes with the main panels.

the correlation time in non-stimulated cells (Supplementary Fig. 4). In agreement with our assumption of linear dynamics<sup>21</sup> and the general prediction of the FRT, we found that the correlation time scales linearly with the response time to small stimulus ( $R^2 = 0.75$ ; Fig. 2b) whereas to large stimulus it scales poorly ( $R^2 = 0.07$ ; Fig. 2b, inset). This result has an important practical implication: The response time that governs the cellular response in chemotaxis can be experimentally inferred by measuring the temporal correlations in behavioural fluctuations from cells before stimulus.

Cellular behavioural variability can also be defined by the amplitude of the noise rather than its temporal correlations. To characterize the amplitude of the output noise of the chemotaxis network, we computed the power spectral density of the switching binary time series measured from individual motors before stimulus (Fig. 3a and Supplementary Fig. 5). We evaluated the low-frequency noise by integrating the power spectrum between  $f = 1/1,500$  s<sup>-1</sup> and  $f = 1/10$  s<sup>-1</sup>. In this frequency range, the temporal fluctuations are putatively caused by the slow methylation–demethylation of the receptor–kinase complexes that are also controlling the adaptive process<sup>14</sup>. Two elements

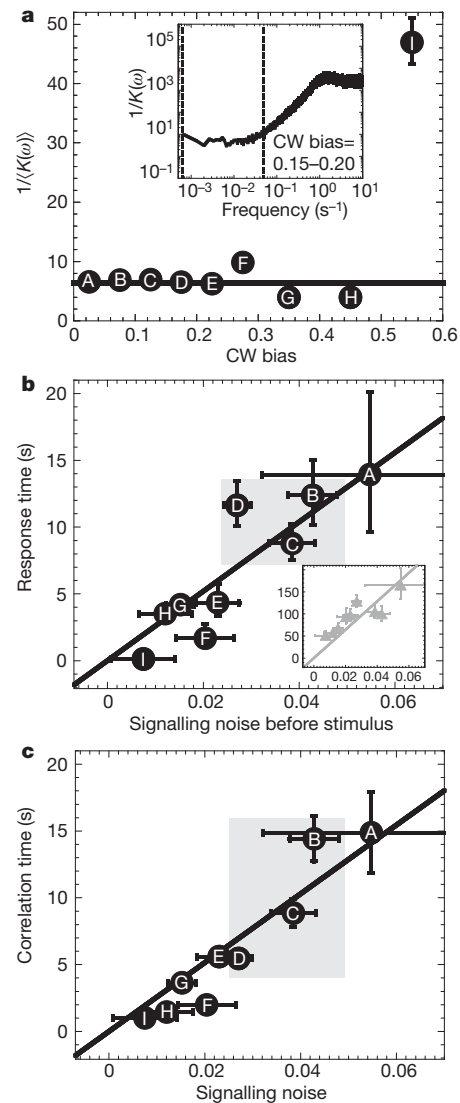




**Figure 3 | Low-frequency noise in non-stimulated cells.** **a**, Low-frequency noise in individual wild-type RP437 cells (black) and RP437 cells expressing CheR from pZE21-CheR (grey) versus CW bias. The inset shows power spectral density as a function of noise frequency. Black line, power density averaged over all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR) with CW bias = 0.15–0.20. Dark grey line, power density of the motor decoupled from the signalling network<sup>3</sup>. We determined the low-frequency noise for the region between the dotted lines. See Supplementary Fig. 5 for all CW bias bins. **b**, Signalling noise as a function of CW bias for wild-type RP437 cells and RP437 cells expressing CheR from pZE21-CheR. Signalling noise is defined as the variance  $\sigma_{\text{CheY-P}}^2$  of the fluctuating [CheY-P]. Letters correspond to the CW bias bins (Fig. 1a). The power spectral densities and CW biases are averaged over cells within the same CW bias. Error bars show the standard error associated with the estimated signalling noise within each bin.

contribute to the observed output noise: the spontaneous noise associated with the signalling events of the chemotaxis network and the stochastic switching behaviour of the bacterial motor (Fig. 3a). The binary nature of the switching behaviour of the motor dominates the variance of the noise and masks the signalling noise within the chemotaxis network the output signalling molecule of which is the phosphorylated form of the signalling protein CheY<sup>1,2</sup>. The active form, CheY-P, binds to the sensory basal part of the flagella rotary motor and induces CW rotation. Using a procedure developed by ref. 22, we decoupled the signalling noise,  $\sigma_{\text{CheY-P}}^2$ , from that of the motor. We then found that the signalling noise decreased with the CW bias (Fig. 3b).

Operationally, we used a simplified expression of the FRT, in which the response function of the chemotaxis system  $\mu(t)$  and the auto-correlation function  $C(t)$  of the spontaneous fluctuations of the cellular behaviour should be related by  $\mu(t) = -K \frac{d}{dt} C(t)$ . Here, the fluctuation-response coupling coefficient  $K$  may depend on the genetic background,



**Figure 4 | Relationship between signalling noise and response time to a small external stimulus.** **a**, Mean coupling coefficient  $1/\langle K(\omega) \rangle$  for each CW bias bin. We computed the geometric mean over frequencies ranging from  $1/1,500 \text{ s}^{-1}$  to  $1/20 \text{ s}^{-1}$ , represented by the dashed lines in the inset to **a**. We found that the coupling coefficient  $K$  for the small stimulus was constant at long timescales for frequencies in this range (see also Supplementary Fig. 7a). The standard error of the mean is smaller than the symbol size except for the highest CW bias bin I. The line is the mean value of  $1/\langle K(\omega) \rangle$  computed over CW biases ranging from 0.00 to 0.5. The inset to **a** shows  $1/K(\omega)$  for cells with a CW bias ranging from 0.15 to 0.20 (10 nM L-aspartate increase). For large stimulus  $K$  is not constant (see Supplementary Fig. 7b). **b**, Average response times of all cells (wild-type RP437 and RP437 expressing inducible CheR) to small stimulus (black circles) or large stimulus (grey triangles in inset to **b**) versus mean pre-stimulus signalling noise. Solid lines are linear fits forced through the origin. Response time =  $C \times \sigma_{\text{CheY-P}}^2$ . Black line:  $C = 259 \pm 25 \text{ s } \mu\text{M}^{-2}$  ( $R^2 = 0.8$ ) for small stimulus. Grey line in inset to **b**:  $C = 3,215 \pm 307 \text{ s } \mu\text{M}^{-2}$  ( $R^2 = 0.4$ ) for large stimulus. Grey area, representative behaviour of a wild-type population. The insets in **b** and **c** show  $1/K(\omega)$  for cells with a CW bias ranging from 0.15 to 0.20 (10 nM L-aspartate increase). **c**, The correlation time as a function of the mean signalling noise before stimulus for all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR). Black line, linear fit function forced through the origin. Correlation time =  $C \times \sigma_{\text{CheY-P}}^2$ .  $C \approx 257 \pm 21 \text{ s } \mu\text{M}^{-2}$  ( $R^2 = 0.9$ ). Letters correspond to the CW bias bins (Fig. 1a). Error bars for the correlation time are the average half-lengths of the first uncorrelated CCW intervals. Error bars for the signalling noise are the standard error associated with the signalling noise in each bin. Grey area is representative behaviour of a wild-type population.

growth conditions, and functional state of the cell. We plotted the coefficient  $K(\omega) = -\frac{2\text{Im}[\tilde{u}(\omega)]}{\omega P(\omega)}$  as a function of CW bias, where  $P(\omega)$  is the power spectral density of the spontaneous fluctuations (Fig. 4a and Supplementary Figs 6 and 7). In the most general non-equilibrium case, the coupling coefficient  $K$  may change when the genetic background or the growth conditions are modified. In chemotaxis, we found that the value of the coupling coefficient  $K(\omega)$  is independent of the functioning states of the cell and levels of expression of the chemotaxis proteins (Fig. 4a). This result is remarkable because most of the chemotaxis network has highly nonlinear signal processing<sup>23,24</sup>.

It is usual to consider that noise is an independent limiting factor in intracellular signalling and that evolution selects network designs to reduce it<sup>25</sup>. However, using the framework of the FRT, we asked whether the temporal fluctuations in the switching rate of the motor and the cellular response are ever dynamically coupled. Remarkably, we found that the response time to a small external stimulus scaled linearly with the signalling noise from the chemotaxis network in cells before stimulus ( $R^2 = 0.8$ ; Fig. 4b), which was consequently linearly related to the correlation time ( $R^2 = 0.9$ ; Fig. 4c). Furthermore, we found that the response time to a large stimulus scaled poorly ( $R^2 = 0.4$ ) with the signalling noise, reflecting that for large stimulus, the system operates outside the regime of linear approximation (Fig. 4b, inset).

We interpret this observation in simple mathematical terms, where the fluctuations in the network output,  $\delta_{\text{CheY-P}}$ , about its average have linearized kinetics in the form of a Langevin equation<sup>21,26</sup>:  $\frac{d}{dt}\delta_{\text{CheY-P}} = -\frac{1}{\tau}\delta_{\text{CheY-P}} + \sqrt{D}\delta\eta(t)$ , where  $\sqrt{D}\delta\eta(t)$  is a white-noise source with intensity  $D$  and  $\tau$  is the measured correlation time in the output of the signalling system. In this coarse-grained picture, there should exist a strict relationship between the signalling output noise amplitude  $\sigma^2_{\text{CheY-P}}$  and the time  $\tau$ , where  $\sigma^2_{\text{CheY-P}} = (D/2)\tau$ . Although the coefficient  $D$  could potentially depend on intracellular parameters in a complex way, our experiments surprisingly showed that two cellular traits,  $\sigma^2_{\text{CheY-P}}$  and the response time, are linearly coupled. This observation implies that the coefficient  $D$  remains approximately constant over a wide range of functioning states of the cell (that is, CW bias). This result is consistent with the fact that the coefficient  $\langle K(\omega) \rangle_\omega$  (Fig. 4a) determines the behaviour of  $D$ , because  $\langle K(\omega) \rangle_\omega \propto 1/D$ . Consequently, we anticipate that below an upper bound imposed mainly by rotational diffusion<sup>27</sup>, cells with the largest behavioural variability before stimulus would also exhibit the strongest chemotactic drift in response to an external stimulus<sup>21</sup>.

Although the FRT predicts the existence of a coupling between cellular response and noise, it does not specify how this coupling depends on the different states of the cell. Therefore, we hypothesize that the specific design of the signalling pathways could govern such interdependence. We find that a simple kinetic model and experimental data support this hypothesis (Supplementary Fig. 8): in chemotaxis, the value of the coefficient  $D$  is governed by the adaptation mechanism that uses the classic futile cycle<sup>21</sup> as a core module in which two antagonistic enzymes regulate the activity of the kinase-receptor complexes. Because the futile cycle is a design shared by a large class of signalling pathways<sup>21,28,29</sup>, it raises the possibility that for these systems, noise and cellular response are coupled in a similar way. To gain general insights into the selection of a specific coupling, we should examine how certain classes of design and function of networks may constrain the behaviour of this interdependence<sup>30</sup>.

## METHODS SUMMARY

**Response time.** For each cell (whose behaviour is defined by a specific CW bias bin), the response time was measured from the time of stimulus through all successive averaged CCW intervals that were longer than the mean pre-stimulus CCW interval length. This mean was obtained by averaging together the CCW

interval lengths chosen at random time points within the binary time series of the non-stimulated cell.

**Correlation time.** To determine the correlation time of the CCW sequences, we used serial correlation coefficients (Supplementary Fig. 4c) for the CCW interval lengths<sup>19,20</sup>. We converted the correlated number of sequences to the real correlation time lengths, including the half-length of the first uncorrelated CCW interval. To determine whether the sequences in each lag (the number of preceding CCW intervals) were correlated, we used the Wilcoxon rank sum test (the “ranksum” Matlab function) at a significance level of  $P = 0.01$  (Supplementary Fig. 4d), as in ref. 20. We considered the first non-zero lag that had  $h = 0$  as the end of the correlation.

**Low-frequency noise and motor noise.** We define the low-frequency noise  $N_i^{\text{LF}}$  of the  $i$ th cell as the integrated power density  $P_i(f)$  of the binary time series from  $f_i = 1/1,500 \text{ s}^{-1}$  to  $f_f = 1/10 \text{ s}^{-1}$ , which is  $N_i^{\text{LF}} \equiv \int_{f_i}^{f_f} P_i(f) df$  (Fig. 3a). We define the low-frequency motor noise  $N_i^{\text{LF,M}}$  as the integrated flat baseline of the power density (Fig. 3a, dark grey line) on the same timescale. We estimated signalling noise from the average experimental power spectral density, the average CW bias, and the gain function between the input signal (steady-state [CheY-P]) and output signal (average CW bias) using methods introduced by ref. 22 (Methods).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 9 June; accepted 4 October 2010.

Published online 14 November 2010.

- Bourret, R. B., Borkovich, K. A. & Simon, M. I. Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annu. Rev. Biochem.* **60**, 401–441 (1991).
- Bourret, R. B. & Stock, A. M. Molecular information processing: lessons from bacterial chemotaxis. *J. Biol. Chem.* **277**, 9625–9628 (2002).
- Korobkova, E. A., Emonet, T., Park, H. & Cluzel, P. Hidden stochastic nature of a single bacterial motor. *Phys. Rev. Lett.* **96**, 058105 (2006).
- Callen, H. B. & Welton, T. A. Irreversibility and generalized noise. *Phys. Rev.* **83**, 34–40 (1951).
- Prost, J., Joanny, J. F. & Parrondo, J. M. Generalized fluctuation-dissipation theorem for steady-state systems. *Phys. Rev. Lett.* **103**, 090601 (2009).
- Bustamante, C., Macosko, J. C. & Wuite, G. J. L. Grabbing the cat by the tail: manipulating molecules one by one. *Nature Rev. Mol. Cell Biol.* **1**, 130–136 (2000).
- Dorignac, J., Kalinowski, A., Erramilli, S. & Mohanty, P. Dynamical response of nanomechanical oscillators in immiscible viscous fluid for *in vitro* biomolecular recognition. *Phys. Rev. Lett.* **96**, 186105 (2006).
- Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
- Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
- Cugliandolo, L. F., Dean, D. S. & Kurchan, J. Fluctuation-dissipation theorems and entropy production in relaxational systems. *Phys. Rev. Lett.* **79**, 2168–2171 (1997).
- Chetrite, R., Falkovich, G. & Gawedzki, K. Fluctuation relations in simple examples of non-equilibrium steady states. *J. Stat. Mech.-Theory E* **2008**, P08005 (2008).
- Speck, T. & Seifert, U. Restoring a fluctuation-dissipation theorem in a nonequilibrium steady state. *Europhys. Lett.* **74**, 391–396 (2006).
- Berg, H. C. Motile behavior of bacteria. *Phys. Today* **53**, 24–29 (2000).
- Korobkova, E., Emonet, T., Vilar, J. M., Shimizu, T. S. & Cluzel, P. From molecular noise to behavioural variability in a single bacterium. *Nature* **428**, 574–578 (2004).
- Cluzel, P., Surette, M. & Leibler, S. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science* **287**, 1652–1655 (2000).
- Sourjik, V. & Berg, H. C. Receptor sensitivity in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **99**, 123–127 (2002).
- Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387**, 913–917 (1997).
- Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).
- Anderson, R. L. Distribution of the serial correlation coefficient. *Ann. Math. Stat.* **13**, 1–13 (1942).
- Ratnam, R. & Nelson, M. E. Nonrenewal statistics of electrosensory afferent spike trains: implications for the detection of weak sensory signals. *J. Neurosci.* **20**, 6672–6683 (2000).
- Emonet, T. & Cluzel, P. Relationship between cellular response and behavioral variability in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **105**, 3304–3309 (2008).
- Shibata, T. & Fujimoto, K. Noisy signal amplification in ultrasensitive signal transduction. *Proc. Natl Acad. Sci. USA* **102**, 331–336 (2005).
- Bray, D., Levin, M. D. & Morton-Firth, C. J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85–88 (1998).
- Sourjik, V. & Berg, H. C. Functional interactions between receptors in bacterial chemotaxis. *Nature* **428**, 437–441 (2004).
- Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237 (2002).

26. Bialek, W. & Setayeshgar, S. Physical limits to biochemical signaling. *Proc. Natl Acad. Sci. USA* **102**, 10040–10045 (2005).
27. Andrews, B. W., Yi, T. M. & Iglesias, P. A. Optimal noise filtering in the chemotactic response of *Escherichia coli*. *PLOS Comput. Biol.* **2**, e154 (2006).
28. Goldbeter, A. & Koshland, D. E. Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci. USA* **78**, 6840–6844 (1981).
29. Detwiler, P. B., Ramanathan, S., Sengupta, A. & Shraiman, B. I. Engineering aspects of enzymatic signal transduction: photoreceptors in the retina. *Biophys. J.* **79**, 2801–2817 (2000).
30. Shinar, G. & Feinberg, M. Structural sources of robustness in biochemical reaction networks. *Science* **327**, 1389–1391 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was funded by an NSF DMR award 0213745 to the Materials Research Science and Engineering Center at the University of Chicago, and NIH award R01AI059195-03 (to P.C.). W.P. and T.E. were supported by NSF CCF0829836, an Alfred P. Sloan Research Fellowship, and a National Academies Keck

Futures Initiative award (to T.E.). J.F.M. was supported by NSF awards PHY-0852130 and DMR-0715099 and NIH grant 1U54CA143869-01. This work was also supported by the Chicago Biomedical Consortium with support from The Searle Funds at The Chicago Community Trust. D. Trentham supplied caged L-aspartate. We thank J. S. Parkinson for  $\Delta$ CheB mutant strains RP4972 and RP4992. We thank T. Shimizu for discussions and sharing unpublished work. We thank H. Lee for help with the HPLC measurements. We thank J. Moffitt and K. Wood for comments on the manuscript and all members of the Cluzel laboratory for many discussions. W. Grus provided editorial assistance.

**Author Contributions** P.C. conceived and designed the research. H.P. performed all the experiments. H.P., P.C., T.E., W.P. and J.F.M. analysed the data. H.P., P.C., J.F.M. and T.E. wrote the paper. C.C.G. constructed the pZE21-CheR plasmid.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to P.C. ([cluzel@mcb.harvard.edu](mailto:cluzel@mcb.harvard.edu)).



## METHODS

**Strains and plasmids.** RP437 is a wild-type *E. coli* strain for chemotaxis<sup>31</sup>. To construct pZE21-CheR, we amplified *cheR* using polymerase chain reaction (PCR) from the chromosome of the RP437 strain with the following primers: CheR-KpnI-5': 5'-gcc ggt acc atg act tca tca tct ctg ccc tg-3' and CheR-HindIII-3': 5'-cgc aag ctt tta atc ctt act tag cgc at-3'. The gene fragment was inserted in the KpnI and HindIII sites of a pZE21 series plasmid<sup>30</sup> that contained a kanamycin resistance cassette and a TetR inducible promoter. The plasmid pZS4-Int1 encodes *tetR* under a constitutive promoter, which modulates the expression of the TetR-regulated *cheR* construct<sup>32</sup>. This plasmid carries a spectinomycin resistance gene. Wild-type cells with and without plasmid exhibited similar noise levels (Fig. 3a) and CCW interval lengths after stimulus (Supplementary Fig. 1a, c and d) at the single-cell level.

**HPLC calibration of the release of aspartate.** We prepared 10  $\mu$ l samples of 0.5-mM caged L-aspartate solution under the same conditions for the chemotaxis experiments and illuminated them with intense ultraviolet light from a Xenon flash lamp (built-in L7685 reflective mirror, 60 W, Hamamatsu). We estimated the relative concentration of the caged L-aspartate in each sample by the high-performance liquid chromatography (HPLC) peak area. By comparing the decreasing HPLC peak area with its initial peak area, we found the released L-aspartate concentration as a function of the number of ultraviolet flashes (Supplementary Fig. 9). The samples released about 1  $\mu$ M L-aspartate per ultraviolet flash. The HPLC gradient conditions had five steps: (1) equilibrium with 20% acetonitrile, 0.1% TFA/80% water, 0.1% TFA; (2) gradient of 20–55% acetonitrile over 30 min; (3) first washing with 55–90% acetonitrile for 20 min; (4) second washing with 90% acetonitrile for 5 min; and (5) equilibrium with 20% acetonitrile, 0.1% TFA/80% water, 0.1% TFA.

**Photo-release and single-cell assay.** We sheared the flagella of the cells by slowly forcing them through a thin needle (inner diameter = 0.19 mm, 27 G  $\frac{1}{2}$ , B-D) 40 times. Cultures grew overnight in 3 ml of tryptone broth at 35 °C with shaking at 200 r.p.m. We transferred the overnight cultures to a 250 ml flask, in which we diluted them 1:50 in 12-ml tryptone broth and grew the cells again at 35 °C at 200 r.p.m. To obtain cells with different CW biases, we induced plasmid expression with various concentrations of anhydrotetracycline (0–2.5 ng ml<sup>-1</sup>) in the diluted overnight cultures. The media also contained the antibiotic specific to the plasmid. We harvested the cells when the absorbance A reached  $\sim 0.3$  at 600 nm. We washed the cells and resuspended them in motility medium (0.1 mM EDTA, 0.1 mM L-methionine, 10 mM potassium phosphate pH 7.0). We prepared glass slides (No. 1  $\frac{1}{2}$ , 18 mm, Corning) coated with poly-L-lysine and a solution of beads (Polybead Amino 1.0  $\mu$ m Microspheres, Polysciences) coated with rabbit antibodies against flagella. We mixed the cells (4–5  $\mu$ l) with the beads (4–5  $\mu$ l) and incubated them for 20 min at room temperature (21–22 °C). This process caused the cell bodies to stick to the glass slide and the beads to attach to the flagella. Although the probability of a bead attaching to a rotating flagellum was low, we consistently obtained a few labelled flagella in each sample. After incubation we removed the unattached cells and beads and then added 8  $\mu$ l of 5  $\mu$ M (for small stimulus) or 500  $\mu$ M (for large stimulus) caged L-aspartate solution to the sample medium. We covered the sample with oil (immersion oil transparent to ultraviolet: type FF, Cargille Laboratories) to prevent evaporation. We placed the sample under a dark-field condenser to produce a bright red image of the bead. Harmful blue light was filtered out by a long-pass filter (NT52-543, Edmund Industrial). We observed the samples under an Olympus IX71 microscope with an oil immersion objective 100 $\times$  (numerical aperture = 1.3, Olympus Uplan FI, oil iris  $\infty/0.17$ ). We recorded the long circular motions of individual beads attached to rotating flagella of single cells through a four-quadrant photomultiplier (type: R5900U-01-M4, Hamamatsu). The signal from the photomultiplier, a four-voltage time series, was monitored with a PC computer via LabView software (National Instrument). The rotation of the bead was simultaneously recorded using a charge-coupled device camera (1/3" midresolution Exview digital B/W camera, Sony). We converted the signal to a binary time series indicating transitions between CCW and CW rotations. After 1,500 s (or 300 s) of recording the rotational motion of the bead, we photo-released the caged aspartate (caged L-aspartic acid, sodium salt (189110): N-[1-(2-nitrophenyl)ethoxycarbonyl]aspartic acid, sodium, C<sub>13</sub>H<sub>13</sub>N<sub>2</sub>O<sub>8</sub>·Na, relative molecular mass 348.2 and molar absorption  $\epsilon = 4,710$  M<sup>-1</sup> cm<sup>-1</sup> at maximum wavelength  $\lambda_{\max} = 264$  nm), from Calbiochem or synthesized by D. Trentham, G. Reid and J. Corrie). We illuminated the sample with an intense ultraviolet light from the Xenon flash coupled into a light guide (A2873, quartz glass fibre, Hamamatsu) and widely focused onto the whole sample with two ultraviolet-coated lenses (focal length = 35 mm and diameter = 25.4 mm; focal length = 20 mm and diameter = 12.7 mm, ThorLabs). These ultraviolet flashes produced a stepwise release of 1  $\mu$ M (or 10 nM) L-aspartate from the 0.5 mM (or 5  $\mu$ M) caged L-aspartate<sup>33</sup>. The magnitude of the stepwise stimulus corresponds to the typical

increase in attractant concentration encountered by bacteria swimming in a gradient of 1 nM  $\mu$ m<sup>-1</sup> (refs 34 and 35).

**Definition of CW bias.** We define  $T_{ij}^{CW}$  and  $T_{ij}^{CCW}$  as the durations of the *j*th CW and CCW intervals of the *i*th cell. The CW bias for the *j*th CW-to-CCW interval pair of the *i*th cell is  $b_{ij} = T_{ij}^{CW} / (T_{ij}^{CCW} + T_{ij}^{CW})$ . The pre-stimulus CW bias of the *i*th cell,  $\langle b_{ij} \rangle_{\text{before}}$ , is the time average of  $b_{ij}$  over a time window of length  $t_{i,\text{before}}$  preceding the stimulus.  $t_{i,\text{before}}$  was 300 s for the cells with CW bias exceeding 0.25 responding to the large stimulus and 1,500 s for all other cells. Similarly, the post-stimulus CW bias of the *i*th cell,  $\langle b_{ij} \rangle_{\text{after}}$ , is the temporal average of  $b_{ij}$  over a time window of duration  $t_{i,\text{after}}$  seconds following the stimulus. For the small (or large) stimulus, the first two (or 200) CW–CCW interval pairs following stimulus were not included.  $t_{i,\text{after}}$  was 1,500 s for small stimuli, 900 s for large stimuli and CW bias < 0.25, and 300 s for large stimuli and CW bias > 0.25.

**Response time.** For each cell (the behaviour of which is defined by a specific CW bias bin), the response time was measured from the time of stimulus through all successive averaged CCW intervals that were longer than the mean pre-stimulus CCW interval length. This mean was obtained by averaging together the CCW interval lengths chosen at random time points within the binary time series of the non-stimulated cell. If the response time included more than one CCW interval, the CW interval length between two successive CCW intervals was also included in the response time. To get the final response time, we subtracted the mean non-stimulated portion of the first responding CCW interval. For example, if the third CCW interval is the last CCW interval length significantly longer than the mean CCW interval length before stimulus (dashed line in Figs 1b and c), the response time would be:

$$\langle T_{CCW, 1st} \rangle + \langle T_{CW, 1st} \rangle + \langle T_{CCW, 2nd} \rangle + \langle T_{CW, 2nd} \rangle + \langle T_{CCW, 3rd} \rangle - \langle T_{CCW, 1st, \text{prestimulus}} \rangle$$

The dashed line in Fig. 1b and c and Supplementary Fig. 1e represents the trend of the mean pre-stimulus CCW interval length in each CW bias bin. Because of the presence of a few outliers, we used the geometric mean to compute the trend of the mean CCW interval lengths after stimulus and mean pre-stimulus CCW interval length within each CW bias bin (Fig. 1b and c).

**Correlation time.** To determine the correlation time of the CCW sequences, we used serial correlation coefficients (Supplementary Fig. 4c) for the CCW interval lengths<sup>19,20</sup>. We converted the correlated number of sequences to the real correlation time lengths, including the half-length of the first uncorrelated CCW interval. To determine whether the sequences in each lag (the number of preceding CCW intervals) were correlated, we used the Wilcoxon rank sum test (the “ranksum” Matlab function) at a significance level of  $P = 0.01$  (Supplementary Fig. 4d) as in ref. 20. We considered the first non-zero lag that had  $h = 0$  as the end of the correlation.

**Low-frequency noise and motor noise.** We define the low frequency noise  $N_i^{LF}$  of the *i*th cell as the integrated power density  $P_i(f)$  of the binary time series from  $f_i = 1/1,500$  s<sup>-1</sup> to  $f_f = 1/10$  s<sup>-1</sup>, which is  $N_i^{LF} \equiv \int_{f_i}^{f_f} P_i(f) df$  (Fig. 3a). We define the low-frequency motor noise  $N_i^{LF,M}$  as the integrated flat baseline of the power density (Fig. 3a, dark grey line) on the same timescale.

**Estimating signalling noise.** To estimate the signalling noise, we used a formula  $\sigma_{M, \text{total}}^2 \cong \sigma_M^2 + g_M^2 \bar{b}^2 \frac{\sigma_{\text{CheY-P}}^2}{[\text{CheY-P}]^2}$  which shows the relationship between the variance  $\sigma_{\text{CheY-P}}^2$  of [CheY-P] and the variance  $\sigma_{M, \text{total}}^2$  of the output signals. This formula was derived from a model recently introduced to describe generally the gain–noise relationship between the input and output signals in the chemical reaction network<sup>22</sup>. As ref. 22 showed, the temporally fluctuating output signal from a well defined steady state (CW bias =  $\bar{b}$ ) due to the fluctuating input signal ([CheY-P]) is described by the following linearized chemical Langevin equation:  $\delta b = \gamma_M \delta[\text{CheY-P}] - \delta b / \tau_M + \xi_M(t)$ , where  $\delta b$  and  $\delta[\text{CheY-P}]$  are small deviations of the CW bias and [CheY-P] from their steady values, respectively,  $\tau_M$  is the typical timescale of the motor alone and  $\xi_M(t)$  is the Gaussian white-noise term that satisfies  $\xi_M(t) = 0$  and  $\xi_M(t) \xi_M(t') = \sigma_{\xi_M}^2 \delta(t - t')$ . From this equation, we obtain the total variance of the output signals due to the temporally fluctuating input signals and the Gaussian white noise:

$$\sigma_{M, \text{total}}^2 = \frac{g_M \bar{b}}{\theta_M} + g_M^2 \bar{b}^2 \frac{\tau_{\text{CheY-P}}}{\tau_M + \tau_{\text{CheY-P}}} \frac{\sigma_{\text{CheY-P}}^2}{[\text{CheY-P}]^2}$$

where  $[\text{CheY-P}]$  is the steady value of fluctuating [CheY-P] values given by:

$$[\text{CheY-P}] = K_M \left( \frac{\bar{b}}{1 - \bar{b}} \right)^{1/N_H}$$

where  $K_M$  (half the concentration of CheY-P that yields CW bias = 0.5) and the Hill coefficient  $N_H$  are given by 3.1  $\mu\text{M}$  and 10.3, respectively, in ref. 15). The constant  $\Theta_M$  in the first term is defined by  $\Theta_M \equiv 2\gamma_M [\text{CheY-P}] / \sigma_{\xi_M}^2$  and  $\bar{b}$  is the CW bias.  $g_M$  is the gain function defined as the ratio of the fractional change of the output signal to the input signal: that is,  $g_M = (\delta b / \bar{b}) / (\delta [\text{CheY-P}] / [\text{CheY-P}]) = N_H (1 - \bar{b})$ , where  $N_H (1 - \bar{b})$  is obtained from ref. 15.  $\tau_{\text{CheY-P}}$  is a characteristic timescale of the  $[\text{CheY-P}]$  fluctuations and is proportional to the input noise  $\sigma_{\text{CheY-P}}^2$  as follows:  $\tau_{\text{CheY-P}} = \frac{\sigma_{\xi_{\text{CheY-P}}}^2}{2} \sigma_{\text{CheY-P}}^2$ . This relationship is derived from the chemical Langevin equation describing the  $[\text{CheY-P}]$  fluctuations from its steady state ( $[\text{CheY-P}]$ ):

$$\dot{\delta[\text{CheY-P}]} = -\frac{\delta[\text{CheY-P}]}{\tau_{\text{CheY-P}}} + \xi_{\text{CheY-P}}(t)$$

where  $\xi_{\text{CheY-P}}(t)$  is a Gaussian white-noise term that satisfies  $\overline{\xi_{\text{CheY-P}}(t)} = 0$  and  $\overline{\xi_{\text{CheY-P}}(t) \cdot \xi_{\text{CheY-P}}(t')} = \sigma_{\xi_{\text{CheY-P}}}^2 \delta(t - t')$ . As long as the external stimulus is small enough, the response time to the stimulus should scale to  $\tau_{\text{CheY-P}}$ . For the broad range of the functioning states of this paper, we have one condition,  $\tau_{\text{CheY-P}} \gg \tau_M$ , in the timescales involved in this system. Under this condition, the above formula for the total variance of the output signals can be simplified to

$$\sigma_{M, \text{total}}^2 \cong \sigma_M^2 + g_M^2 \bar{b}^2 \frac{\sigma_{\text{CheY-P}}^2}{[\text{CheY-P}]^2}$$

where  $\sigma_{M, \text{total}}^2$  is given by  $\bar{b}(1 - \bar{b})$  for any binary time series and is equal to the integral of the power spectral density over all frequencies (black line in Supplementary Fig. 5) averaged over all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR) and  $\sigma_M^2$  is equal to the integral of the power density (dark grey line in Supplementary Fig. 5) of the isolated motor. We approximated the baseline of the motor power density by finding the mean value of the flat regime (from  $f_i = 1/10 \text{ s}^{-1}$  to  $f_f = 1/5 \text{ s}^{-1}$ ) of the average experimental power density and extending the baseline to the lowest frequency. By using the simplified formula above, we estimated the  $\sigma_{\text{CheY-P}}^2$  values in each CW bias bin (Fig. 3b).

**Definition of noise.** We hypothesize that a small number of proteins and thermally activated biochemical reaction rates cause stochastic fluctuations between functional states of signalling proteins. Operationally, we monitor the cellular behaviour in a motility medium that does not support growth but allows bacteria to perform chemotaxis. Under these conditions, the observed noise does not result from protein synthesis or degradation; rather, it results from fluctuations in protein functional states about a well-defined steady state.

31. Parkinson, J. S. & Houts, S. E. Isolation and behavior of *Escherichia coli* deletion mutants lacking chemotaxis functions. *J. Bacteriol.* **151**, 106–113 (1982).
32. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/11-12 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
33. Jasuja, R., Yu-Lin, Trentham, D. R. & Khan, S. Response tuning in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **96**, 11346–11351 (1999).
34. Adler, J. A method for measuring chemotaxis and use of the method to determine optimum conditions for chemotaxis by *Escherichia coli*. *J. Gen. Microbiol.* **74**, 77–91 (1973).
35. Bainer, R., Park, H. & Cluzel, P. A high-throughput capillary assay for bacterial chemotaxis. *J. Microbiol. Methods* **55**, 315–319 (2003).

# Cap binding and immune evasion revealed by Lassa nucleoprotein structure

Xiaoxuan Qi<sup>1</sup>, Shuiyun Lan<sup>2</sup>, Wenjian Wang<sup>3</sup>, Lisa McLay Schelde<sup>2</sup>, Haohao Dong<sup>1</sup>, Gregor D. Wallat<sup>1</sup>, Hinh Ly<sup>2</sup>, Yuying Liang<sup>2</sup> & Changjiang Dong<sup>1</sup>

**Lassa virus, the causative agent of Lassa fever, causes thousands of deaths annually and is a biological threat agent, for which there is no vaccine and limited therapy. The nucleoprotein (NP) of Lassa virus has essential roles in viral RNA synthesis and immune suppression, the molecular mechanisms of which are poorly understood. Here we report the crystal structure of Lassa virus NP at 1.80 Å resolution, which reveals amino (N)- and carboxy (C)-terminal domains with structures unlike any of the reported viral NPs. The N domain folds into a novel structure with a deep cavity for binding the m7GpppN cap structure that is required for viral RNA transcription, whereas the C domain contains 3'-5' exoribonuclease activity involved in suppressing interferon induction. To our knowledge this is the first X-ray crystal structure solved for an arenaviral NP, which reveals its unexpected functions and indicates unique mechanisms in cap binding and immune evasion. These findings provide great potential for vaccine and drug development.**

Several arenaviruses, including Lassa virus (LASV), can cause severe viral haemorrhagic fevers in humans with high morbidity and mortality, to which there is no vaccine and limited treatment<sup>1–4</sup>. These pathogenic arenaviruses are public health threats and potential biological threat agents. LASV, like other arenaviruses, is a single-stranded ambisense RNA virus with two genomic RNA segments encoding four genes<sup>1</sup>. The NP encapsidates viral genomic RNAs into ribonucleoprotein (RNP) complexes and is required for both RNA replication and transcription<sup>5–7</sup>. Like bunyaviruses and orthomyxoviruses, arenaviruses snatch the cap structure of cellular mRNAs to use as primers to initiate viral transcription, the exact mechanism of which is unknown. The cap-snatching mechanism of arenaviruses seems to be unique, as evidenced by the cytoplasmic localization and the much shorter 5' non-templated mRNA sequences<sup>8–11</sup>. Severe arenavirus infections including lethal Lassa cases are associated with a generalized immune suppression in the infected hosts<sup>12–18</sup>, the exact mechanism of which is unclear but is thought to involve NP's ability to suppress the induction of type I interferon (IFN)<sup>19,20</sup>. To address the functional mechanisms of NP in viral RNA synthesis and host immune suppression, we set out to determine the crystal structure for LASV NP, knowledge derived from which can be extended to other arenavirus NP proteins, as all known arenaviral NP proteins share high sequence identity (Supplementary Fig. 1).

## Structure determination

The full-length 569-residue LASV NP protein (Josiah strain) was expressed and purified as a recombinant MBP fusion protein in *Escherichia coli* as described in Methods. The purified protein exists mainly in two forms, with a majority in trimeric and some in hexameric form. Both forms bind random RNAs, which are longer and more abundant in the hexamers than in the trimers, a feature that is similar to known NPs from negative-strand RNA viruses<sup>21–24</sup>. We attempted to crystallize both forms, but only the trimeric NP formed crystals. The crystals showed heavy twinning with a twin fraction of ~0.43 and the reflection intensity statistic  $|E^2 - 1|$  0.681/0.681. Initial phases were obtained in a space group of *P*321 using the multiple wavelength

anomalous diffraction (MAD) with Samarium derivative. The true space group was *P*3 with three subunits in an asymmetric unit. The structure was refined to a resolution of 1.80 Å with de-twinning. The crystal structures do not contain RNA, indicating that only RNA-free NP was able to form crystals. The final structural model of the native LASV NP has an  $R_{\text{factor}}$  of 0.18 and an  $R_{\text{free}}$  of 0.20. Data collection, phasing and refinement statistics are provided in Supplementary Table 1.

## Overall structure of LASV NP protein

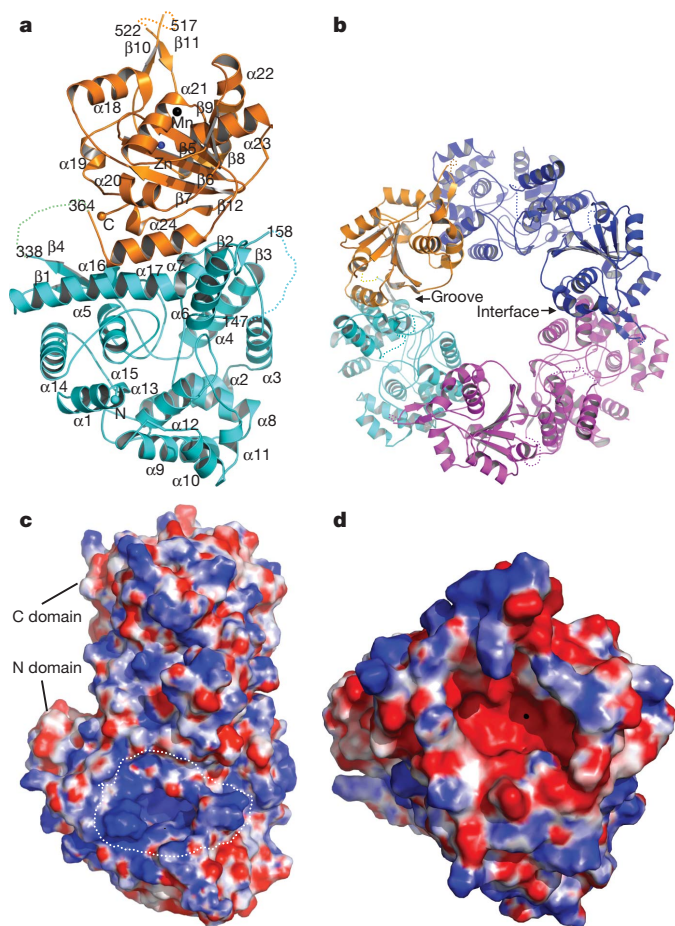
In the NP protomer structure, 514 residues of the 569-residue LASV NP protein were built into the model (Fig. 1a). The electron densities for residues 1–6, 147–157, 339–363, 518–521, 562–569 were not well defined. LASV NP protomer, like other viral NPs<sup>7–10</sup>, is composed of the N- and the C-terminal domains, but neither domain shows structural similarity to any known viral NPs (Supplementary Table 2). The large N domain (residues 7–338) consists mainly of  $\alpha$ -helices and coils, whereas the C domain (residues 364–561) forms a typical  $\alpha/\beta/\alpha$  sandwich architecture (Supplementary Text 1). In the trimeric form, three subunits lie in a head-to-tail orientation to form a ring-shaped structure with a three-fold symmetry (Fig. 1b and Supplementary Fig. 2). Surface rendering reveals a deep cavity located near the bottom of the N domain and a large cavity at the top of the C domain (Fig. 1c, d), which are the cap-binding site and the 3'-5' exoribonuclease active site (see below), respectively. The interface area between the subunits is 455 Å<sup>2</sup>, representing 1.9% of total surface area of a subunit (23,343 Å<sup>2</sup>). The central hole of the trimeric structure is 23 Å in diameter, whereas the head ring is 98 Å and the body ring is 118 Å (Supplementary Fig. 2).

## LASV NP is a 3'-5' exoribonuclease

A Dali search ([http://ekhidna.biocenter.helsinki.fi/dali\\_server](http://ekhidna.biocenter.helsinki.fi/dali_server)) identified several structures similar to the C domain of NP, including several known 3'-5' exonucleases/exoribonucleases in bacteria and humans (for example, human TREX1) (Supplementary Text 2), all of which belong to the DEDDH subfamily of the DEDD (DnaQ) superfamily<sup>25–27</sup>. The

<sup>1</sup>Biomedical Sciences Research Complex, School of Chemistry, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, UK. <sup>2</sup>Department of Pathology and Laboratory Medicine, Emory University School of Medicine, 615 Michael St, Atlanta, Georgia 30322, USA. <sup>3</sup>Laboratory of Department of Surgery, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan Road II, Guangzhou, Guangdong 510080, China.

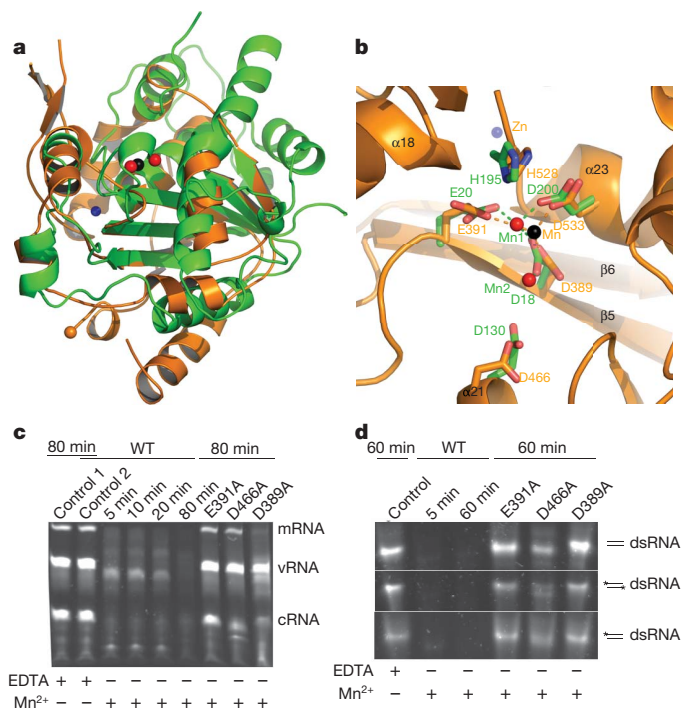




**Figure 1 | The crystal structure of LASV NP protein.** **a**, Cartoon diagram of the LASV NP protomer. The N domain is in cyan with the cyan sphere indicating the N terminus; the C domain is in orange with the orange sphere indicating the C terminus. The black sphere shows  $Mn^{2+}$ , whereas the blue sphere shows  $Zn^{2+}$ . The dotted lines represent the disordered loops. **b**, The ring-shaped structure of LASV NP trimer. The first protomer is coloured as in **a**, the second protomer is in blue and the third is in magenta. The groove and the interface are indicated by arrows. **c**, Electrostatic surface potential map of the NP protomer. The entrance of the cap-binding cavity is shown as a white dotted circle. The blue area represents positively charged residues and the red area represents negatively charged residues. **d**, Electrostatic surface potential map of the 3'-5' exonuclease cavity. The black sphere represents  $Mn^{2+}$ .

human TREX1 structure shows two  $Mn^{2+}$  cations in the active site<sup>27</sup>. We identified one  $Mn^{2+}$  in each subunit of LASV NP by crystal fluorescent scanning, but could not identify the second  $Mn^{2+}$ , possibly because it was not well ordered in the absence of the RNA substrate. The C domain of NP superimposes well with the portion of TREX1 that coordinates the  $Mn^{2+}$  cations (Fig. 2a), in particular the  $\beta 5$ ,  $\beta 6$ ,  $\beta 7$ ,  $\beta 8$  and  $\beta 9$  strands of NP completely overlap with the central  $\beta$ -sheets of TREX1. The putative exonuclease catalytic residues D389, E391, D466, D533 and H528 are absolutely conserved in all known arenavirus NP proteins and are located at identical positions as in the TREX1 active cavity (Fig. 2b). Taken together, the structural evidence indicates that LASV NP is a new member of the DEDD 3'-5' exonuclease superfamily.

We conducted *in vitro* assays to characterize the 3'-5' exonuclease activity of the wild-type LASV NP, as well as NP mutants at putative catalytic sites. We showed that the wild-type protein, in its trimeric or hexameric form, could digest both DNA and RNA substrates (Supplementary Figs 3 and 4). As divalent cations are essential for exonuclease activity<sup>25</sup>, we determined what divalent cation was most effective for NP exonuclease to digest various single-stranded RNA (ssRNA) species that are based on the NP gene in the viral genomic sense



**Figure 2 | The C domain of LASV NP is a 3'-5' exonuclease.**

**a**, Superimposition of the C domain (orange) with human TREX1 protein (green) reveals a high degree of similarity between the two structures.  $Mn^{2+}$  is in black in LASV NP and red in TREX1;  $Zn^{2+}$  is in blue. **b**, The exonuclease catalytic residues of LASV NP and TREX1 are located in identical positions, and are shown in orange for NP and green for TREX1. **c**, The exonuclease activities of the wild-type (WT) and mutant LASV NP with different ssRNAs as substrates. Control 1 contains 10 mM EDTA and no NP. Control 2 contains 10 mM EDTA and NP. **d**, Comparison of the wild-type and NP catalytic mutants in degrading the dsRNA substrates, the 5'-hydroxyl dsRNA (top), double 5'-triphosphorylated dsRNA (middle), and the single 5'-triphosphorylated dsRNA (bottom).

(60 nucleotides, vRNA), complementary antigenomic sense (30 nucleotides, cRNA), or in capped mRNA form (126 nucleotides, mRNA) (Methods). We showed the order of efficiency as  $Mn^{2+} > Co^{2+} > Mg^{2+} > Ca^{2+} > Zn^{2+} > Fe^{2+} > Ni^{2+} > Cu^{2+}$  (Supplementary Fig. 5). Wild-type NP could cleave various ssRNA species efficiently (Fig. 2c), regardless of whether they contained a hydroxyl (5'OH) group, triphosphate (5'ppp), or a cap at the 5' termini (Methods). In contrast, the NP catalytic mutants (D389A, E391A and D466A) showed markedly reduced RNase activity (Fig. 2c and Supplementary Fig. 4). In addition, we showed that wild-type NP, but not its catalytic mutants, could digest cellular RNA substrates *in vitro* with a preference towards short RNA species over long ones (for example, 18s rRNA versus  $\beta$ -globin mRNA, the large versus small fragments in the RNA ladder) (Supplementary Figs 6 and 7). We also demonstrated that wild-type NP, but not its catalytic mutants (D389A, E391A and D466A), can efficiently degrade various dsRNA molecules with 5'-hydroxyl (5'OH), single 5'-triphosphorylate (5'ppp/5'OH) and double 5'-triphosphorylate (5'ppp/5'ppp), as well as the long dsRNA mimic poly(I:C) (Fig. 2d and Supplementary Fig. 8).

Fluorescence scanning analysis identified a zinc ion in the NP structure, despite the fact that no typical zinc finger motif was predicted from the amino acid sequence and that no zinc compounds were used during the purification and crystallization processes. Although the residues C506, C529, H509 and E399 that coordinate the zinc ion are not of the typical zinc-binding motif<sup>28</sup>, they appear to adopt a zinc finger fold in structure<sup>28,29</sup>. The CCHE zinc-binding site is located in the C domain near the 3'-5' exonuclease active site (Supplementary Fig. 9). We speculate that zinc binding may be required to

stabilize the structure of the C domain and/or contribute to the substrate binding and specificity of the exonuclease activity<sup>28,29</sup>. A highly positively charged groove located between the N and C domains is predicted as the genomic RNA-binding site (Supplementary Fig. 10). An *in vitro* assay confirmed that RNAs are bound within the purified NP oligomers and protected from its intrinsic exonuclease activity (Supplementary Figs 10 and 11, Supplementary Text 3 and Methods).

### Exonuclease and immune evasion

To determine whether the exoribonuclease activity is important for the transcriptional function of NP, we generated alanine substitution at five putative catalytic sites, D389A, E391A, D466A, D533A and H528A, in the mammalian cell expression vectors of either native or Myc-tagged NP gene, and examined the activity of each mutant in transcribing the LASV minigenome RNA that encodes a *Renilla* luciferase (RLuc) reporter gene<sup>20</sup> (Methods). As shown in Fig. 3a, each NP mutant expressed comparable protein levels to the wild type, and led to similar folds of increase in RLuc activity, indicating that these mutations did not alter the overall structure (Supplementary Text 4 and Supplementary Fig. 12) or affect the basic function of NP in mediating viral RNA transcription.

We next examined whether the exoribonuclease activity is required for NP's function in the suppression of IFN<sup>19,20</sup>. As expected, wild-type NP strongly inhibited Sendai-virus-induced IFN- $\beta$  activation by a promoter assay (Methods), whereas all the catalytic mutants D389A, E391A, D466A, D533A and H528A showed a complete loss of function at a low level of transfected expression vectors (10 ng) and showed various levels of deficiency at higher levels (Fig. 3b and

Supplementary Fig. 13). Our results confirm a previous study showing that the D389 residue of LASV NP, as well as its corresponding residue D382 in the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV), is required for IFN suppression but not for viral RNA transcription<sup>30</sup>, and may help to explain the loss of IFN suppression for Tacaribe virus NP (Supplementary Fig. 14 and Supplementary Text 5). In summary, these data provide strong genetic evidence for an important role of the NP exoribonuclease activity in suppressing the IFN induction.

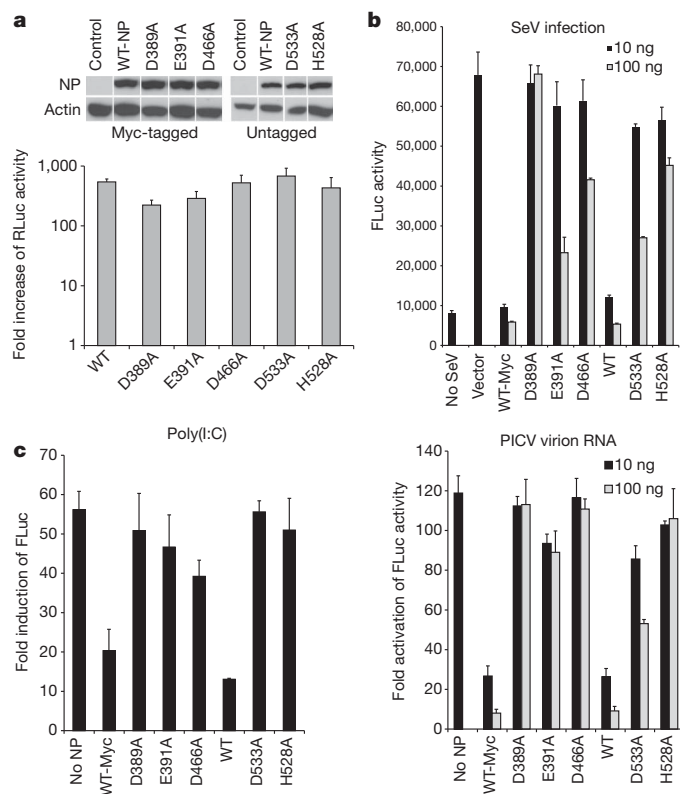
Viral infections are usually detected by the cellular pattern-recognition receptors (PRRs) such as toll-like receptors (TLRs) and cytosolic RNA sensors, retinoid-acid-inducible gene-I-like helicase (RIG-I) and melanoma differentiation-associated protein 5 (MDA5), which recognize the pathogen-associated molecular patterns (PAMP) RNA ligands and initiate signalling pathways to induce the production of type I IFNs<sup>31,32</sup>. We hypothesize that NP prevents the virus-induced IFN induction by degrading the PAMP RNA ligands that otherwise would trigger the viral sensors in the cells.

We examined whether the NP RNase function is essential for suppressing the IFN production induced by the immunostimulatory RNAs, that is, poly(I:C) and the virion RNAs extracted from Pichinde virus, which is a prototypic arenavirus<sup>33</sup>. We found that whereas wild-type NP efficiently inhibited the IFN- $\beta$  activation induced by poly(I:C) or by Pichinde-virion-associated RNAs, none of the five catalytic mutants (D389A, E391A, D466A, D533A and H528A) exhibited any suppressive activity (Fig. 3c). Similar results had been reported for LCMV NP<sup>34</sup>.

We have shown that the NP exoribonuclease activity is essential for suppressing both viral-infection-induced and immunostimulatory-RNA-induced IFN production. A good example of exonuclease-mediated suppression of IFN production has been demonstrated for human TREX1 protein, which degrades small ssDNAs and dsDNAs accumulated during cellular apoptosis. Failure to clear these DNA fragments by TREX1 natural mutants leads to the activation of cellular DNA receptors to trigger a persistent production of IFNs that contributes to human autoimmune diseases<sup>27,35–38</sup>. How does the NP RNase activity function in suppressing the virus-induced IFN production? A simplistic but reasonable model is that the NP RNase activity is able to remove viral PAMP RNAs that are otherwise recognized by the cellular PRRs. Although we have shown that LASV NP protein can degrade various RNA templates *in vitro*, we believe that the NP RNase activity must be highly regulated *in vivo*, as NP does not cause a generalized nonspecific RNA degradation process of cellular or viral RNAs in the cells (Supplementary Text 6 and Supplementary Fig. 15). We propose that the NP RNase activity in the cells is restricted to viral PAMP RNAs through a yet-to-be characterized regulatory mechanism. A recent publication has shown a direct protein–protein interaction of NP with RIG-I and MDA5 (ref. 34), which may be one possible mechanism for the specific nuclease activity of NP against these PRR-associated PAMP RNAs.

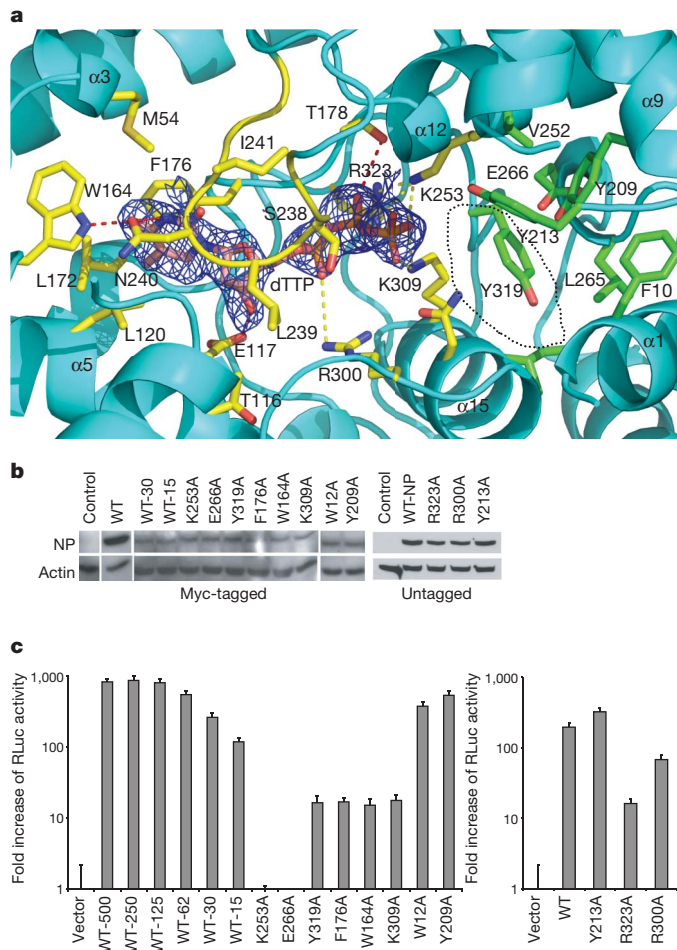
### LASV NP is a cap-binding protein

The N domain adopts a completely novel fold not found in the Dali server. To identify the cap-binding residues in the deep cavity of the N domain, we attempted to soak and perform co-crystallization of LASV NP with m7GpppG, triphosphorylated, diphosphorylated or monophosphorylated ribonucleotides (Methods). We could observe the clear density for the triphosphate and partial density for uridine (Supplementary Fig. 16) from the triphosphorylated ribonucleotide complex structures. We also visualized the structure of NP in complex with dTTP with a clear original  $F_o - F_c$  electron density contoured at  $2.5\sigma$  for dTTP (Fig. 4a). The triphosphate group of dTTP was bound in the middle of the cavity in an identical manner as that of UTP (Supplementary Fig. 16), in which it was anchored by salt bonds formed with the side chains of the conserved residues K309, R300, R323 and K253. In the deep end of the cavity, thymidine occupied a hydrophobic pocket that is composed of residues F176, W164, L172,



**Figure 3 | The exonuclease activity of NP is important for blocking the IFN induction.** Results shown are the average ( $n = 3$ ) with error bars indicating the standard deviations. **a**, The NP catalytic mutants were expressed at similar levels to the wild type in mammalian cells and had similar transcriptional activities in the LASV minigenome assay. **b**, The NP catalytic mutants were defective in suppressing the Sendai-virus (SeV)-induced IFN induction by a LUC-based IFN- $\beta$  promoter assay. **c**, The NP catalytic mutants were defective in suppressing the IFN production induced by the immunostimulatory RNAs poly(I:C) and Pichinde-virion-associated RNAs.





**Figure 4 | The cap-binding residues and their roles in viral RNA transcription.** Results shown are the average ( $n = 3$ ) with error bars indicating the standard deviations. **a**, A cap analogue dTTP is bound within the deep cavity of the N domain of LASV NP. Original  $F_o - F_c$  map for the dTTP in blue contoured at  $2.5\sigma$ . The F176 and W164 or L172 (L120) residues form a typical cap-binding sandwich structure. The middle cavity binds the triphosphate moiety and the hydrophobic cavity entrance can accommodate the second base of the cap structure. The carbon atoms are in pink for the dTTP, in yellow for the deep cavity residues and in green for the cavity entrance residues. **b**, The NP mutants were expressed at similar levels as the wild type at 15–30 ng plasmid (WT-15, WT-30) in the transfected mammalian cells. **c**, Mutational analyses of the residues within the cap-binding cavity for the transcriptional activity using the LASV minigenome assay.

M54, L120, L239 and I241. We propose that this dTTP-binding pocket is the binding site for the cap structure m7GTP and that the residues located within the pocket may have to change conformation to accommodate the cap moiety. Although the N domain of NP is not structurally similar to any of the cap-binding proteins (Supplementary Table 3), its hydrophobic thymidine-binding pocket shares common features for cap binding<sup>39,40</sup>. Moreover, the NP cap-binding cavity has a unique feature in that its entrance contains another hydrophobic region that is composed of the hydrophobic residues Y319, Y209, Y213, L265 and the acidic residue E266, which can potentially act as the binding site for the second base of the m7GpppN (where N represents G, C, U or A) cap structure. The entrance of the cap-binding cavity has an oval shape with a diameter of 9–13 Å, which is a perfect fit for the single-stranded mRNA. We propose that a loop composed of residues K236 to S242 serves as a ‘gate’ for the capped template (primer) binding and that the entire structure of m7GpppN, including the cap m7G, the triphosphate, and at least one more nucleotide, is embedded within the deep cavity. This binding feature is unlike other known cap-binding proteins, in which only the m7G caps are locked in

between the sandwich, whereas the rest of the RNA molecule is exposed<sup>39,40</sup>.

To characterize the role of the cap-binding residues in viral RNA transcription, we examined a panel of NP mutants with alanine substitution of residues located inside and at the entrance of the cavity and that are conserved among all known arenaviruses for their ability to mediate the cap-dependent viral RNA transcription using the LASV minigenome replicon assay (Methods). Wild-type NP (with or without Myc tag) produced up to a 1,000-fold increase in RLuc reporter activity over a control reaction, and more than 100-fold increase even when expressed at a low level (15 ng of transfected NP plasmid DNAs). All mutant proteins were expressed at similar levels as the wild type transfected with 15–30 ng of plasmid (Fig. 4b). Compared to the wild type, the K253A and E266A mutants completely lost the RNA transcription activity, and the Y319A, F176A, W164A, K309A and R323A mutants showed significantly decreased activity (Fig. 4c). R300A had a minor effect, whereas W12A and Y209A had no effect. It is worth noting that none of these mutants was found to impact the NP function in the suppression of IFN (Supplementary Fig. 17). These functional data correlate well with the proposed cap-binding function of some of these conserved residues.

The unique cap-binding feature of LASV NP, in that the entire cap structure m7GpppN is buried within the cavity, has significant implications in understanding the distinctive cap-snatching mechanism of arenaviruses. Once NP binds and protects the 5' cap m7GpppN, the rest of the mRNA molecule located outside of the cavity may be susceptible to viral and/or host exonuclease-mediated degradation and/or to endonuclease-mediated cleavage (Supplementary Fig. 9). This may help to explain the relatively short (1–4 nucleotides) 5' non-templated sequences in arenavirus mRNAs<sup>1,8,9</sup>. However, individual mutation of the NP exonuclease catalytic sites did not show any defect in viral cap-dependent RNA transcription (Fig. 3a), indicating that the NP exonuclease activity is not essential (required) for generating the capped primers. It is worth noting that we did not identify an influenza polymerase PA-like endonuclease structural motif<sup>41,42</sup> within LASV NP structure (Supplementary Table 4). Instead, recent studies indicated that the LASV L polymerase protein contains an endonuclease domain in its N terminus that is crucial for the cap-dependent viral RNA transcription<sup>43,44</sup>.

## Conclusion

Our structural analysis and functional assays have demonstrated that the C domain of LASV NP contains 3'–5' exoribonuclease activity that is required for suppressing IFN- $\beta$  induction. We have provided evidence to suggest that the NP RNase activity is highly regulated in cells and proposed a novel mechanism by which the NP RNase activity may specifically remove the viral PAMP RNA ligands to suppress the production of IFN. Another important feature of LASV NP protein is that its N domain contains a deep cavity to bind and shield the entire m7GpppN cap structure, which is distinct from other known cap-binding proteins, and has shed light on the unique cap-snatching mechanism of arenaviruses. In addition, we have also identified an unusual zinc-binding site and the viral RNA-binding groove in the LASV NP structure. Taken together, these findings reveal several new and potentially vulnerable targets on NP for the development of antivirals and effective vaccines to combat LASV and other pathogenic arenaviruses that can cause severe haemorrhagic fever diseases in humans.

## METHODS SUMMARY

The crystals were grown using the sitting-drop technique, and the native structure was determined with the MAD data. All the NP mutations were generated using the QuikChange site-directed mutagenesis kit (Stratagene) and confirmed by DNA sequencing. The RNA synthesis assays used the LASV minigenome (MG) system, and the Sendai-virus-induced IFN- $\beta$  activation assay was conducted as described<sup>45</sup>. The immunostimulatory RNA-induced IFN- $\beta$  activation assay was



conducted by transfecting HEK293 cells with the IFN- $\beta$ -LUC promoter construct and either wild-type or mutant NP construct, followed by Lipofectamine-2000-mediated transfection of poly(I:C) or Pichinde-virion-isolated RNAs. Activation of the IFN- $\beta$  promoter was quantified by measuring the LUC activity.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 7 May; accepted 25 October 2010.**

**Published online 17 November 2010.**

- Buchmeier, M. J., De La Torre, J. C. & Peters, C. J. in *Fields Virology* Vol. 2 (eds Knipe, D. M. & Howley, P. M.) 1791–1827 (Lippincott Williams & Wilkins, 2007).
- Delgado, S. *et al.* Chapare virus, a newly discovered arenavirus isolated from a fatal hemorrhagic fever case in Bolivia. *PLoS Pathog.* **4**, e1000047 (2008).
- Briese, T. *et al.* Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.* **5**, e1000455 (2009).
- Khan, S. H. *et al.* New opportunities for field research on the pathogenesis and treatment of Lassa fever. *Antiviral Res.* **78**, 103–115 (2008).
- Hass, M., Golnitz, U., Muller, S., Becker-Ziaja, B. & Gunther, S. Replicon system for Lassa virus. *J. Virol.* **78**, 13793–13803 (2004).
- Pinschewer, D. D., Perez, M. & de la Torre, J. C. Role of the virus nucleoprotein in the regulation of lymphocytic choriomeningitis virus transcription and RNA replication. *J. Virol.* **77**, 3882–3887 (2003).
- Lopez, N., Jacamo, R. & Franze-Fernandez, M. T. Transcription and RNA replication of tacaribe virus genome and antigenome analogs require N and L proteins: Z protein is an inhibitor of these processes. *J. Virol.* **75**, 12241–12251 (2001).
- Polyak, S. J., Zheng, S. & Harnish, D. G. 5' termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. *J. Virol.* **69**, 3211–3215 (1995).
- Meyer, B. J. & Southern, P. J. Concurrent sequence analysis of 5' and 3' RNA termini by intramolecular circularization reveals 5' nontemplated bases and 3' terminal heterogeneity for lymphocytic choriomeningitis virus mRNAs. *J. Virol.* **67**, 2621–2627 (1993).
- Jin, H. & Elliott, R. M. Characterization of Bunyamwera virus S RNA that is transcribed and replicated by the L protein expressed from recombinant vaccinia virus. *J. Virol.* **67**, 1396–1404 (1993).
- Plotch, S. J., Bouloy, M., Ulman, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**, 847–858 (1981).
- Edington, G. M. & White, H. A. The pathology of Lassa fever. *Trans. R. Soc. Trop. Med. Hyg.* **66**, 381–389 (1972).
- Fisher-Hoch, S., McCormick, J. B., Sasso, D. & Craven, R. B. Hematologic dysfunction in Lassa fever. *J. Med. Virol.* **26**, 127–135 (1988).
- McCormick, J. B. & Fisher-Hoch, S. P. Lassa fever. *Curr. Top. Microbiol. Immunol.* **262**, 75–109 (2002).
- Baize, S. *et al.* Early and strong immune responses are associated with control of viral replication and recovery in lassa virus-infected cynomolgus monkeys. *J. Virol.* **83**, 5890–5903 (2009).
- Baize, S. *et al.* Lassa virus infection of human dendritic cells and macrophages is productive but fails to activate cells. *J. Immunol.* **172**, 2861–2869 (2004).
- Mahanty, S. *et al.* Cutting edge: impairment of dendritic cells and adaptive immunity by Ebola and Lassa viruses. *J. Immunol.* **170**, 2797–2801 (2003).
- Muller, S., Geffers, R. & Gunther, S. Analysis of gene expression in Lassa virus-infected HuH-7 cells. *J. Gen. Virol.* **88**, 1568–1575 (2007).
- Martinez-Sobrido, L., Giannakas, P., Cubitt, B., Garcia-Sastre, A. & de la Torre, J. C. Differential inhibition of type I interferon induction by arenavirus nucleoproteins. *J. Virol.* **81**, 12696–12703 (2007).
- Martinez-Sobrido, L., Zuniga, E. I., Rosario, D., Garcia-Sastre, A. & de la Torre, J. C. Inhibition of the type I interferon response by the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. *J. Virol.* **80**, 9192–9199 (2006).
- Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**, 357–360 (2006).
- Tawar, R. G. *et al.* Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* **326**, 1279–1283 (2009).
- Ye, Q., Krug, R. M. & Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078–1082 (2006).
- Albertini, A. A. *et al.* Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**, 360–363 (2006).
- Cisneros, G. A. *et al.* Reaction mechanism of the epsilon subunit of *E. coli* DNA polymerase III: insights into active site metal coordination and catalytically significant residues. *J. Am. Chem. Soc.* **131**, 1550–1556 (2009).
- Zuo, Y. *et al.* Crystal structure of RNase T, an exonuclease involved in tRNA maturation and end turnover. *Structure* **15**, 417–428 (2007).
- de Silva, U. *et al.* The crystal structure of TREX1 explains the 3' nucleotide specificity and reveals a polyproline II helix for protein partnering. *J. Biol. Chem.* **282**, 10537–10543 (2007).
- Hall, T. M. Multiple modes of RNA recognition by zinc finger proteins. *Curr. Opin. Struct. Biol.* **15**, 367–373 (2005).
- Matthews, J. M. & Sunde, M. Zinc fingers—folds for many occasions. *IUBMB Life* **54**, 351–355 (2002).
- Martinez-Sobrido, L. *et al.* Identification of amino acid residues critical for the anti-interferon activity of the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. *J. Virol.* **83**, 11330–11340 (2009).
- Kawai, T. & Akira, S. Innate immune recognition of viral infection. *Nature Immunol.* **7**, 131–137 (2006).
- McCartney, S. A. & Colonna, M. Viral sensors: diversity in pathogen recognition. *Immunol. Rev.* **227**, 87–94 (2009).
- Lan, S. *et al.* Development of infectious clones for virulent and avirulent pichinde viruses: a model virus to study arenavirus-induced hemorrhagic fevers. *J. Virol.* **83**, 6357–6362 (2009).
- Zhou, S. *et al.* Induction and inhibition of type I interferon responses by distinct components of lymphocytic choriomeningitis virus. *J. Virol.* **84**, 9452–9462 (2010).
- Crow, Y. J. & Rehwinkel, J. Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. *Hum. Mol. Genet.* **18**, R130–R136 (2009).
- Lee-Kirsch, M. A. *et al.* Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nature Genet.* **39**, 1065–1067 (2007).
- Lehtinen, D. A., Harvey, S., Mulcahy, M. J., Hollis, T. & Perrino, F. W. The TREX1 double-stranded DNA degradation activity is defective in dominant mutations associated with autoimmune disease. *J. Biol. Chem.* **283**, 31649–31656 (2008).
- Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**, 587–598 (2008).
- Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15**, 500–506 (2008).
- Fechter, P. & Brownlee, G. G. Recognition of mRNA cap structures by viral and cellular proteins. *J. Gen. Virol.* **86**, 1239–1249 (2005).
- Yuan, P. *et al.* Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* **458**, 909–913 (2009).
- Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**, 914–918 (2009).
- Leike, M., Brunotte, L., Busch, C. & Gunther, S. An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome. *J. Virol.* **84**, 1934–1944 (2010).
- Morin, B. *et al.* The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. *PLoS Pathog.* **6**, e1001038 (2010).
- Lan, S., McLay, L., Aronson, J., Ly, H. & Liang, Y. Genome comparison of virulent and avirulent strains of the Pichinde arenavirus. *Arch. Virol.* **153**, 1241–1250 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** C.D. wishes to thank J. Naismith for his continuing support, encouragement and advice; R. M. Elliott for providing the human  $\beta$ -globin containing plasmid pHRL-CMV and his critical reading of the manuscript; H. Liu for discussions; and L. Major for the pLou3 plasmid. C.D. is a Wellcome trust career development fellow (083501/Z/07/Z). We thank the staff at IO2 and IO3 beam stations of Diamond light sources for their assistance with data collection. This work was supported in part by funds from the Southeast Regional Center of Excellence for Emerging Infections and Biodefense (5-U54-AI-057157-06), the pilot component of the U19 grant (5-U19-AI057266-07), and the Emory University Research Committee (URC) to Y.L. and H.L.; a seed grant from the Emory DDRDC (DK64399) and a research scholar grant from the American Cancer Society (RSG-06-162-01-GMC) to H.L.; NIH grant R01AI083409 to Y.L.; and NIH grant AI067704 to T. G. Parslow and Y.L. We thank K. Curtis and T. W. Geisbert for providing us with non-infectious genomic RNA samples of LASV, and J. Aronson for the PICV virus.

**Author Contributions** X.Q., S.L., W.W., L.M.S., H.D. and G.D.W. performed experiments. C.D., Y.L. and H.L. conceived the idea for the study, performed some of the assays, participated in the analysis and interpretation of the data, and wrote the manuscript. All authors contributed to the final version of the manuscript.

**Author Information** Atomic coordinates and structure factors for the structures have been deposited in the Protein Data Bank under accession codes 3MWP for the native, 3MWT for the manganese complex, 3MX2 for the dTTP complex and 3MX5 for the UTP complex. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to C.D. (cd26@st-andrews.ac.uk), Y.L. (yliang5@emory.edu) or H.L. (hly@emory.edu).

## METHODS

**Protein expression and purification.** The full-length LASV NP gene (Josiah strain) was cloned into the pMAL-c2X-derived pLou3 plasmid, downstream of the TEV cleavage site following the MBP gene. This construct, encoding the N-terminal MBP tagged NP protein, was transformed into Rosetta cells (Novagen). After IPTG induction at a final concentration of 0.03 mM overnight at 20 °C, the cells were harvested by centrifugation at 8,000 r.p.m. for 20 min and suspended in TEN buffer (20 mM Tris, pH 7.5; 0.2 M NaCl, 10% glycerol, 1 mM EDTA) with protease inhibitors (Roche), 1  $\mu$ M DNase (Sigma) and 1 mM phenylmethylsulphonyl fluoride (Sigma). After cells were lysed by a cell disruptor (Constant System Ltd), the cell lysates were collected by centrifugation at 20,000 r.p.m. for 30 min and applied on an amylose column. The column was washed with >10-column volumes of the sample buffer. The MBP–NP fusion protein was eluted with the TEN buffer containing 10 mM maltose. The MBP–NP fusion protein was then cleaved by Tev proteinase. The MBP portion was removed through two amylose columns, and the NP protein was purified to homogeneity by gel filtration column. Trypsin digestion coupled with mass spectroscopy confirmed that the purified LASV NP protein was homogenous (data not shown), with a final concentration of 7 mg ml<sup>-1</sup>.

**Crystallization and data collection.** A Cartesian robot (Genomic solutions) was used to screen for optimal crystallization conditions. The native crystals were obtained in 0.2 M LiCl<sub>2</sub> and 20% PEG3350 in 1 week at 20 °C. To obtain the NP complex with m7GpppG, m7GTP, or m7GDP, the NP protein was incubated with individual compound at a concentration of 2 mM for 30 min on ice and the crystallization conditions were screened. The NP complex with other triphosphorylated, diphosphorylated or monophosphorylated nucleotides were formed by incubating the NP protein with 50 mM of the respective compounds for 30 min on ice and the crystallization conditions were screened. The crystallization conditions were optimized until the resolution of the data was better than 2.5 Å. All crystals grew in 0.2 M KCl or 0.2 M LiCl and 14–22% PEG3350. The NP complexed with manganese ion was obtained by crystallizing the NP protein in 0.2 M MnCl<sub>2</sub>, 25% PEG3350 followed by soaking the crystals in 0.2 M NaCl<sub>2</sub>, 20% PEG3350 and 15% glycerol three times for 15 min each. The presence of the manganese and the zinc ions was confirmed in all the crystals by fluorescence scanning at the Diamond light sources UK. All the crystals were protected by cryoprotectants that contain 15% to 20% glycerol in the crystallization conditions before data collection in IO2 or IO3 at the Diamond light sources UK. The Samarium derivative crystals were obtained by soaking the crystals overnight in 100 mM Samarium acetate, 0.2 M LiCl and 16% PEG3350, and was protected in a cryoprotectant of 0.2 M LiCl, 16% PEG3350 and 20% glycerol. The Samarium derivative MAD data were collected at a wavelength of 1.83 Å for peak data, 1.84 Å for inflection data and 1.45 Å for remote data from a single crystal. All the data were indexed, integrated and scaled by HKL2000 or Mosflm and Scale. **Structure determination.** The crystals were heavily twinned with a twinning fraction of 0.43. The initial phases were obtained from a space group of P321 using the MAD data and SOLVE<sup>46</sup>. The initial model was built using RESOLVE<sup>46</sup>, Buccaneer and Coot<sup>47</sup>. It was found that the true space group of the crystals was P3 during the structure refinement. The structures were refined using REFMAC5<sup>48</sup>, and the water molecules were added into the structure by ARP/wARP<sup>49</sup>. The  $F_o - F_c$  maps for ligands (dTTP, UTP, zinc and manganese) were calculated before any ligand was added into the structures. The structures were de-twinned at last using REFMAC5, and the structures were evaluated using Molprobit<sup>50</sup>.

**In vitro RNA synthesis.** The 30-nucleotide cRNA (sense) sequence 5'-CUGGGC UUACCUAUUCUCAGCUGAUGACCC-3' was derived from the LASV NP (Josiah strain) S segment (nucleotides 2186–2215 in antigenomic orientation) and chemically synthesized by Eurogentec. The 30-nucleotide vRNA (in genomic orientation) sequence 5'-GGGUCAUCAGCUGAGAAUAGGUAAGCCAG-3' was complementary to the cRNA. The cRNA (30 nucleotides) was used as one of the three substrates for 3'-5' exoribonuclease assay. To obtain the blunted dsRNA, both cRNA and vRNA oligonucleotides were dissolved into 0.1 M NaCl, 1 mM EDTA and 0.1 M Tris pH 8.0 at the final concentration of 200 mM, and an equal amount of the two oligonucleotides was mixed together and annealed in a thermocycler as follows: 95 °C for 3 min, 68 °C for 1 min and then 4 °C.

The 5'-triphosphorylated vRNA was generated by *in vitro* transcription of the partial dsDNA template formed by the T7 promoter sequence 5'-AATTTAA TACGACTCACTATAGG-3' and the reverse complement of the T7 promoter sequence and of the LASV (Josiah strain) S segment (nucleotides 2186–2215) 5'-CTGGGCTTACCTATTCTCAGCTGATGACCCTATAGTGAGTCGTATT AAATT-3' using the T7 MEGAshortscript kit following the manufacturer's instructions (Ambion). A similar strategy was used to generate the 32-nucleotide triphosphorylated cRNA with the T7 primer and LASV (Josiah strain) S segment (nucleotides 2186–2213) 5'-GGGTCATCAGCTGAGAATAGGTAAGCCCA GCCTATAGTGAGTCGTATTAAATT-3'. A similar strategy was used to generate

the 60-nucleotide vRNA corresponding to LASV (Josiah strain) S segment (nucleotides 2186–2213), using the partial dsDNA template formed by 5'-AATTTAA TACGACTCACTATAGG-3' and 5'-GTAAATCCCTGCAGTCGGCAGGGTTTA CCGCTGGGCTTACCTATTCTCAGCTGATGACCCTATAGTGAGTCGTATT AAATT-3' as a template. To generate the doubly 5'-triphosphorylated dsRNA, equal amounts of the triphosphorylated 5'ppp-vRNA and 5'ppp-cRNA were annealed *in vitro*. To make the singly 5'-triphosphorylated dsRNA, equal amounts of the *in vitro* synthesized 32-nucleotide 5'ppp-vRNA and the chemically synthesized 30-nucleotide unphosphorylated cRNA were annealed *in vitro*. The human 18S rRNA fragment (128 nucleotides) was generated by a T7 RNA polymerase-directed *in vitro* RNA synthesis reaction, using the pTRI-RNA 18S control plasmid (Ambion), following the manufacturer's instruction.

To synthesize the capped viral mRNA transcripts corresponding to nucleotides 992–1117 of the LASV NP gene, the DNA template was PCR amplified from the NP expression plasmid with a forward primer 5'-AATTTAATACGACTCAC TATAGGGAAACACTGTCGTTGATCTGGAATC-3' (underlined are T7 promoter sequences) and a reverse primer 5'-GGGTCATCAGCTGAGAATAG GTAAGCCCCAGCGG-3', and subjected to *in vitro* RNA synthesis using the mMESSAGE mMACHINE T7 Ultra kit (Ambion) following the manufacturer's instruction, except that no poly(A) tail was added.

A plasmid pHRL-CMV that encodes the T7 promoter (T7p)-directed human  $\beta$ -globin gene was provided by R. Elliott and G. Blakqori. The T7p-globin DNA fragment was purified by agarose electrophoresis after digestion of the pHRL-CMV plasmid with HindIII and SmaI. The capped human globin mRNA transcripts were generated using the T7p-globin fragment as a template and the mMESSAGE mMACHINE T7 Ultra kit from Ambion, and the poly(A) tail was added following the manufacturer's instruction.

The ssRNA markers (perfect RNA markers, 0.1–1 kb) were purchased from Novagen. The low molecular mass ssRNA marker (10–100 nucleotides) was purchased from USB. The dsRNA ladder (21–500 bp) was purchased from New England Biolabs.

**In vitro 3'-5' exoribonuclease assays.** The *in vitro* 3'-5' exoribonuclease assays were carried out in 10  $\mu$ l of the reaction solution containing 0.3 M NaCl, 10% glycerol, 20 mM Tris pH 7.5, 10 mM MnCl<sub>2</sub>, 7  $\mu$ g of either wild-type or mutant NP proteins, and 8 units of the RNaseIN inhibitor (Promega), in the presence of various substrate(s), at 37 °C for 60–100 min. The control reactions included all but MnCl<sub>2</sub>, which was substituted by 20 mM EDTA. All the reactions, each in triplicate, were stopped by the addition of EDTA to a final concentration of 20 mM. The samples were mixed with equal volumes of RNA loading buffer (Ambion), heated at 95 °C for 3 min, cooled on ice for 5 min, and separated in 15% or 6% urea-polyacrylamide gel, or 2% agarose gel. The gels were stained in 0.05% ethidium bromide for 25 min, visualized using the 2UV transilluminator (UVP).

**The luciferase-based assay to quantify virus-induced and immunostimulatory RNA-induced interferon- $\beta$  activation.** The Sendai-virus-induced IFN- $\beta$  activation assay was conducted as described previously<sup>45</sup>. In brief, 293T cells were co-transfected using calcium phosphate with 100 ng of a vector that expresses the firefly luciferase (FLuc) reporter gene from a known functional promoter sequence of the IFN- $\beta$  gene (pIFN $\beta$ -LUC), variable amounts of either wild-type or mutant LASV NP vectors, and 50 ng of a  $\beta$ -gal-expressing plasmid for transfection normalization. At 24 h after transfection, cells were infected with Sendai virus (at multiplicity of infection = 1) to induce IFN- $\beta$  expression. At 24 h after infection, cell lysates were prepared for luciferase and  $\beta$ -gal assays. FLuc activities were normalized by the  $\beta$ -gal values. Each transfection was conducted in triplicate and repeated in at least two independent experiments.

To determine whether NP can suppress the immunostimulatory RNA-induced IFN production, HEK293 cells were transfected with pIFN $\beta$ -LUC, variable amounts of either wild-type or mutant LASV NP vectors, and a  $\beta$ -gal-expressing plasmid for transfection normalization. Eighteen hours later, cells were transfected with either 1  $\mu$ g of poly(I:C) or 250 ng of Pichinde virion RNA by lipofectamine 2000. Luciferase activity was determined at 18 h after the immunostimulatory RNA transfection and normalized by the  $\beta$ -gal activity.

**Pichinde virion RNA preparation.** Pichinde viruses were purified by 20% sucrose gradient ultracentrifugation at 50,000g for 2 h. Virus RNA was extracted with RNABee (Tel Test) according to the manufacturer's protocol.

**LASV minigenome (MG) transcription assay.** The full-length LASV L and NP genes (Josiah strain) were cloned into the pCAGGS vector for expression in mammalian cells. The LASV MG construct contains the T7 promoter-directed LASV S-segment-like sequences that include all the important cis-acting elements required for viral RNA synthesis (5' UTR, intergenic region and 3' UTR) and encode a *Renilla* luciferase (RLuc) gene in place of the viral NP coding sequence. This LASV-based LUC-encoding minigenome (MG) RNA was transcribed *in vitro* by the T7 MEGAscript kit (Ambion) and transfected into 293T cells, together with the LASV L expression plasmid, and wild-type or mutant NP expression plasmid. A  $\beta$ -gal expression vector was included in each

transfection to normalize for cell transfection efficiency. LUC activity was determined at 24 h after transfection, normalized by  $\beta$ -gal activity, and shown as fold increase over a control sample that lacked the L expression plasmid. Each reaction was conducted in triplicate and in at least two independent experiments.

46. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
48. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
49. Cohen, S. X. *et al.* Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D* **60**, 2222–2229 (2004).
50. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).



# Trapped antihydrogen

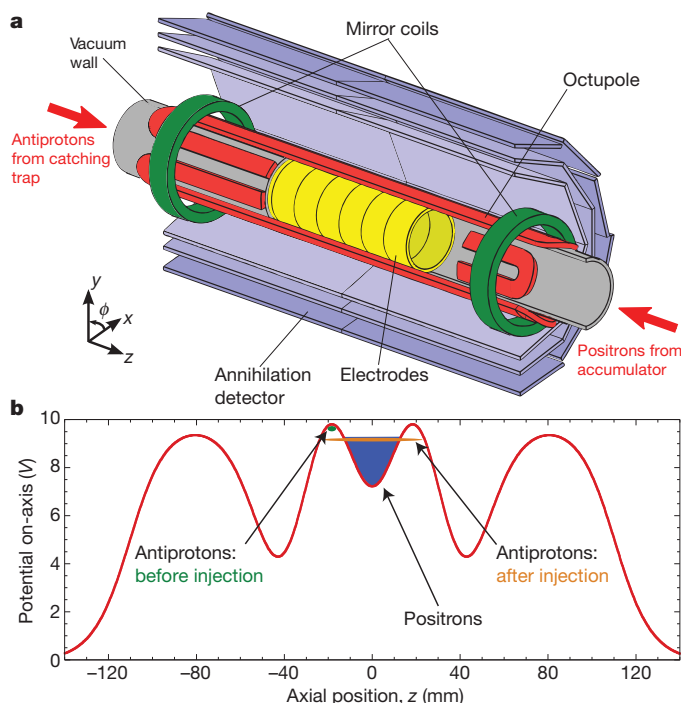
G. B. Andresen<sup>1</sup>, M. D. Ashkezari<sup>2</sup>, M. Baquero-Ruiz<sup>3</sup>, W. Bertsche<sup>4</sup>, P. D. Bowe<sup>1</sup>, E. Butler<sup>4</sup>, C. L. Cesar<sup>5</sup>, S. Chapman<sup>3</sup>, M. Charlton<sup>4</sup>, A. Deller<sup>4</sup>, S. Eriksson<sup>4</sup>, J. Fajans<sup>3,6</sup>, T. Friesen<sup>7</sup>, M. C. Fujiwara<sup>8,7</sup>, D. R. Gill<sup>8</sup>, A. Gutierrez<sup>9</sup>, J. S. Hangst<sup>1</sup>, W. N. Hardy<sup>9</sup>, M. E. Hayden<sup>2</sup>, A. J. Humphries<sup>4</sup>, R. Hydomako<sup>7</sup>, M. J. Jenkins<sup>4</sup>, S. Jonsell<sup>10</sup>, L. V. Jørgensen<sup>4</sup>, L. Kurchaninov<sup>8</sup>, N. Madsen<sup>4</sup>, S. Menary<sup>11</sup>, P. Nolan<sup>12</sup>, K. Olchanski<sup>8</sup>, A. Olin<sup>8</sup>, A. Povilus<sup>3</sup>, P. Pusa<sup>12</sup>, F. Robicheaux<sup>13</sup>, E. Sarid<sup>14</sup>, S. Seif el Nasr<sup>9</sup>, D. M. Silveira<sup>15</sup>, C. So<sup>3</sup>, J. W. Storey<sup>8†</sup>, R. I. Thompson<sup>7</sup>, D. P. van der Werf<sup>4</sup>, J. S. Wurtele<sup>3,6</sup> & Y. Yamazaki<sup>15,16</sup>

Antimatter was first predicted<sup>1</sup> in 1931, by Dirac. Work with high-energy antiparticles is now commonplace, and anti-electrons are used regularly in the medical technique of positron emission tomography scanning. Antihydrogen, the bound state of an antiproton and a positron, has been produced<sup>2,3</sup> at low energies at CERN (the European Organization for Nuclear Research) since 2002. Antihydrogen is of interest for use in a precision test of nature's fundamental symmetries. The charge conjugation/parity/time reversal (CPT) theorem, a crucial part of the foundation of the standard model of elementary particles and interactions, demands that hydrogen and antihydrogen have the same spectrum. Given the current experimental precision of measurements on the hydrogen atom (about two parts in  $10^{14}$  for the frequency of the 1s-to-2s transition<sup>4</sup>), subjecting antihydrogen to rigorous spectroscopic examination would constitute a compelling, model-independent test of CPT. Antihydrogen could also be used to study the gravitational behaviour of antimatter<sup>5</sup>. However, so far experiments have produced antihydrogen that is not confined, precluding detailed study of its structure. Here we demonstrate trapping of antihydrogen atoms. From the interaction of about  $10^7$  antiprotons and  $7 \times 10^8$  positrons, we observed 38 annihilation events consistent with the controlled release of trapped antihydrogen from our magnetic trap; the measured background is  $1.4 \pm 1.4$  events. This result opens the door to precision measurements on anti-atoms, which can soon be subjected to the same techniques as developed for hydrogen.

Charged particles of antimatter can be trapped in a high-vacuum environment in Penning–Malmberg traps, which use axial electric fields generated by hollow cylindrical electrodes and a solenoidal magnetic field to provide confinement. The ALPHA apparatus, located at the Antiproton Decelerator<sup>6</sup> at CERN, uses several such traps to accumulate, cool and mix charged plasmas of antiprotons and positrons to synthesize antihydrogen atoms at cryogenic temperatures. ALPHA evolved from the ATHENA experiment, which demonstrated production and detection of cold antihydrogen at CERN in 2002<sup>2</sup>.

In addition to the charged particle traps necessary to produce antihydrogen, ALPHA features a novel, superconducting magnetic trap<sup>7</sup> (Fig. 1) designed to confine neutral antihydrogen atoms through interaction with their magnetic moments. The atom trap—a variation on the Ioffe–Pritchard minimum-magnetic-field geometry<sup>8</sup>—comprises a transverse octupole<sup>9,10</sup> and two solenoidal ‘mirror’ coils, and surrounds the interaction region where antihydrogen atoms are produced. In comparison with a quadrupole field (used in traditional atom traps) producing an equal trap depth, the transverse field of an

octupole has been shown to greatly reduce the perturbations on charged plasmas<sup>9,10</sup>. The liquid helium cryostat for the magnets also cools the vacuum wall and the Penning trap electrodes; the latter are measured to be at about 9 K. Antihydrogen atoms that are formed with low enough kinetic energy can remain confined in the magnetic trap, rather than annihilating on the Penning electrodes. The ALPHA trap can confine ground-state antihydrogen atoms with a kinetic energy, in



**Figure 1 | The ALPHA central apparatus and mixing potential.**

**a**, Antihydrogen synthesis and trapping region of the ALPHA apparatus. The atom-trap magnets, the modular annihilation detector and some of the Penning trap electrodes are shown. An external solenoid (not shown) provides a 1-T magnetic field for the Penning trap. The drawing is not to scale. The inner diameter of the Penning trap electrodes is 44.5 mm and the minimum-magnetic-field trap has an effective length of 274 mm. Each silicon module is a double-sided, segmented silicon wafer with strip pitches of 0.9 mm in the  $z$  direction and 0.23 mm in the  $\phi$  direction. **b**, The nested-well potential used to mix positrons and antiprotons. The blue shading represents the approximate space charge potential of the positron cloud. The  $z$  position is measured relative to the centre of the atom trap.

<sup>1</sup>Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. <sup>2</sup>Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. <sup>3</sup>Department of Physics, University of California, Berkeley, California 94720-7300, USA. <sup>4</sup>Department of Physics, Swansea University, Swansea SA2 8PP, UK. <sup>5</sup>Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-972, Brazil. <sup>6</sup>Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>7</sup>Department of Physics and Astronomy, University of Calgary, Calgary, Alberta T2N 1N4, Canada. <sup>8</sup>TRIUMF, 4004 Wesbrook Mall, Vancouver, British Columbia V6T 2A3, Canada. <sup>9</sup>Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. <sup>10</sup>Fysikum, Stockholm University, SE-10691, Stockholm, Sweden. <sup>11</sup>Department of Physics and Astronomy, York University, Toronto, Ontario M3J 1P3, Canada. <sup>12</sup>Department of Physics, University of Liverpool, Liverpool L69 7ZE, UK. <sup>13</sup>Department of Physics, Auburn University, Auburn, Alabama 36849-5311, USA. <sup>14</sup>Department of Physics, Nuclear Research Center NEGEV, Beer Sheva, IL-84190, Israel. <sup>15</sup>Atomic Physics Laboratory, RIKEN, Saitama 351-0198, Japan. <sup>16</sup>Graduate School of Arts and Sciences, University of Tokyo, Tokyo 153-8902, Japan. <sup>†</sup>Present address: Physik-Institut, Zürich University, CH-8057 Zürich, Switzerland.

temperature units, of less than about 0.5 K. The extreme experimental challenges are to synthesize such cold atoms from plasmas of charged particles whose electrostatic potential energies can be of order 10 eV—or  $10^5$  K—and to unequivocally identify rare occurrences of trapped antihydrogen against background processes.

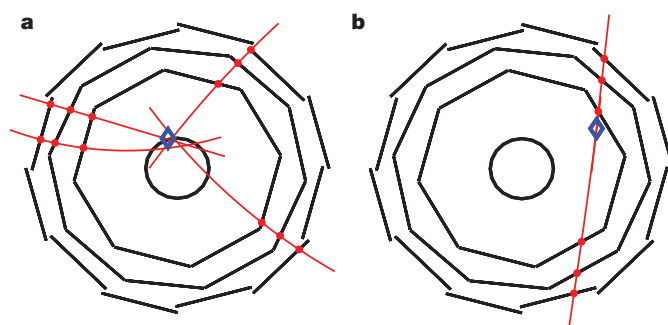
The ALPHA apparatus is designed to demonstrate antihydrogen trapping by releasing the magnetically trapped anti-atoms and detecting their annihilations. A key feature of the device is the ability to turn off the magnetic trapping fields with a time constant of about 9 ms, which is a response several orders of magnitude faster than in typical superconducting systems. Another essential component of ALPHA is an imaging, three-layer, silicon vertex detector<sup>11</sup> (Fig. 1), which is used to identify and locate antiproton annihilations from released antihydrogen atoms and to reject background from cosmic rays that happen to arrive during the time window of interest, when the trap is being de-energized. The magnets have a unique, low-density construction<sup>7</sup> to minimize scattering of annihilation products (pions) so that the positions ('vertices') of antiproton annihilations can be accurately determined.

A trapping attempt involves first preparing clouds of antiprotons and positrons for 'mixing' to produce antihydrogen. The antiproton cloud contains about 30,000 particles obtained from one extracted bunch ( $\sim 3 \times 10^7$  particles at 5.3 MeV) from the Antiproton Decelerator. The antiprotons are slowed in a thin foil, dynamically trapped<sup>12</sup> in a 3-T Penning trap (the 'catching' trap, not shown in Fig. 1) with 3.4-keV well depth, cooled using electrons<sup>13</sup> and then separated from the electrons using pulsed electric fields. The resulting plasma has a radius of 0.8 mm, a temperature of about 200 K and a density of  $6.5 \times 10^6 \text{ cm}^{-3}$ . The positrons are supplied by a  $^{22}\text{Na}$  radioactive source and a Surko-type accumulator<sup>14,15</sup>. To increase the antihydrogen formation rate and trapping probability, the positrons transferred from the accumulator are evaporatively cooled<sup>16,17</sup> (Methods) to about 40 K. The resulting positron plasma has  $2 \times 10^6$  particles, a radius of 0.9 mm and a density of  $5.5 \times 10^7 \text{ cm}^{-3}$ .

Antiprotons and positrons are made to interact within a nested-well axial potential<sup>18</sup> (Fig. 1b) at the centre of the magnetic atom trap. After the two species are placed in their respective potential wells, the superconducting magnets of the atom trap are ramped up to their maximum fields in 25 s. The antiprotons are then excited into the positron plasma using an oscillating electric field that autoresonantly<sup>19,20</sup> increases their energy (Methods). This novel technique is essential for introducing the antiprotons into the positron cloud at low relative velocity, so that antihydrogen can be formed with low energy, and to reduce the heating of the positron plasma.

The positrons and antiprotons interact for 1 s to produce antihydrogen before the uncombined charged particles are ejected from the trap volume. During this mixing time, we record  $5,000 \pm 400$  triggers in the silicon detector. The detector is triggered when charged particles (principally pions) from an antiproton annihilation deposit energy (above a threshold value) in at least two of the inner silicon modules. Cosmic rays can also trigger the detector and do so at a measured rate of  $10.49 \pm 0.03 \text{ Hz}$ . Each trigger can initiate a read-out of position information for the entire detector; the maximum read-out rate for such 'events' is 500 Hz. The position information can be analysed to identify pion trajectories (tracks) to locate antiproton annihilation vertices. An antiproton annihilation can usually be distinguished from a cosmic ray by considering their respective track topologies; see examples in Fig. 2. The rate at which we detect cosmic rays that could be misidentified as antiproton annihilations is  $(4.6 \pm 0.1) \times 10^{-2} \text{ Hz}$  (Methods). Using the spatial distribution of the reconstructed annihilations during mixing<sup>21</sup>, we infer that about 70% of the mixing events are due to impacts from antihydrogen atoms that are not trapped; the remaining ones are mostly antiprotons from atoms that are sufficiently weakly bound to be field-ionized by Penning trap electric fields before reaching the wall.

The magnetic gradients of the atom trap can also act to trap bare antiprotons. Such 'mirror-trapped' antiprotons could escape and annihilate



**Figure 2 | Detected antiproton annihilation and cosmic ray events.**

**a, b,** Projected end views ( $x$ - $y$  plane) of an antiproton annihilation (**a**) and a cosmic-ray event (**b**) detected by the ALPHA detector. The reconstruction algorithm identifies the antiproton vertex (blue diamond) near the Penning trap wall (black circle). The high-energy cosmic ray passes in a near-straight line through the detector, and the vertex-finding algorithm attempts to identify it as a two-track annihilation with an unphysical vertex.

when the magnetic trap is de-energized, mimicking the sought-after signal of trapped antihydrogen atoms being released. After the 1-s mixing period, the charged particles in the mixing trap wells are ejected from the experiment. We then apply four pulses of axial electric 'clearing' fields of up to  $500 \text{ V m}^{-1}$  to remove mirror-trapped antiprotons. The manipulations after mixing take 172 ms, after which we initiate the trap shutdown. The rapid turn-off causes the superconducting elements to 'quench', or become normally conducting. We look for antiproton annihilations from released antihydrogen in a time window of 30 ms (more than three e-folding times for the confining fields) after the start of the magnet shutdown.

We conducted the above-described search experiment 335 times, in three variations. In one variation, referred to as 'left bias' (101 attempts), we erect a static electric field just before the quench to deflect any remaining antiprotons to the left (negative  $z$  direction) of the apparatus as they are released. The second variation, 'right bias' (97 attempts), features a static electric field that should deflect antiprotons to the other side of the device. In the third variation, 'no-bias' (137 attempts), all electrodes are at ground during the magnet quench. The bias electric field has a strength of about  $500 \text{ V m}^{-1}$ . The use of bias fields allows us to use the annihilation imaging detector to distinguish between the release of trapped antihydrogen—which is neutral and is therefore unaffected by these fields—and that of mirror-trapped antiprotons.

To ensure that any detected events are in fact antihydrogen and to eliminate other sources of background, we repeated the above experiments using heated positrons. Following the method introduced by the ATHENA<sup>2</sup> collaboration, we heat the positrons (without particle loss) to about 1,100 K by driving their axial motion. The effect in ALPHA is twofold: antihydrogen formation is suppressed because of the temperature dependence of the three-body process that dominates this reaction<sup>22</sup>, and any antihydrogen formed is unlikely to be trapped because the antiprotons approach thermal equilibrium with the hot positrons through Coulomb collisions. The number of annihilation events during the 1-s mixing time with heated positrons is  $97 \pm 16$ . Apart from the heating of the positrons, the experimental trapping sequence is identical to that described above.

Table 1 summarizes the results of all trapping and background attempts. In the total sample of attempts (335) with cold positrons, we observe 38 annihilations, for a rate of 0.11 events per attempt. For the background sample with heated positrons, we observe one annihilation in 246 attempts, or a rate of 0.0041 events per attempt.

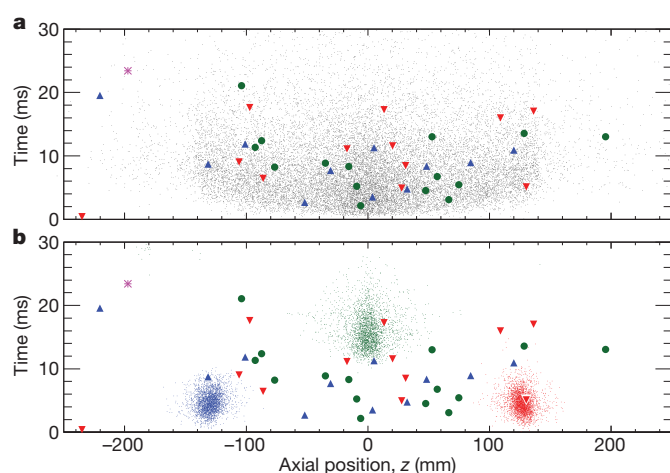
The discrimination provided by the silicon detector and the fast shutdown of our magnetic trap render the cosmic background negligible in comparison with the signal level in the current work. In the integrated observation time ( $335 \times 30 \text{ ms}$ ), we would expect  $0.46 \pm 0.01$  counts to result from misidentified cosmic rays.

**Table 1 | Number of annihilations identified in the 30 ms following the trap shutdown**

Type of attempt	Number of attempts	Antiproton annihilation events
No bias	137	15
Left bias	101	11
Right bias	97	12
No bias, heated positrons	132	1
Left bias, heated positrons	60	0
Right bias, heated positrons	54	0

We consider the effect of the bias fields in Fig. 3. We plot the event time versus the  $z$  coordinate of the reconstructed vertex for all identified annihilations in the 30-ms window. The start of the magnet shutdown corresponds to the zero of time. Figure 3a shows the  $t$ - $z$  distribution for the 38 annihilations recorded using cold positrons and the one annihilation from heated positrons. Superimposed is a scatter plot from a dynamical simulation that predicts the behaviour of trapped antihydrogen atoms being released and annihilating on the Penning trap electrodes. (Details of the simulation procedures are given in Methods.) Figure 3b compares the measured annihilation distribution with simulations of mirror-trapped antiprotons released during the magnet shutdown. Predictions for the left-, right- and no-bias variations are shown.

Particles can be mirror-trapped when the ratio of their transverse to longitudinal energies exceeds a threshold determined by the field geometry. Although the phase space distribution of hypothetical mirror-trapped antiprotons is unknown, we illustrate here the prediction for an initial sample of antiprotons that has a uniform spatial distribution and a flat velocity distribution up to a maximum kinetic energy of 75 eV. This choice is quite conservative, as the maximum longitudinal potential well depth during the mixing process is less than 21 eV. We note that the model predicts that only mirror-trapped antiprotons with a transverse kinetic energy of greater than 45 eV could remain trapped after the clearing pulses. We have not been able to identify any mechanism that could create such antiprotons in the course of our experimental procedure, much less one that would then fail to create them when the positrons are heated by only 0.1 eV.



**Figure 3 | Distributions of released antihydrogen atoms and antiprotons.** **a**, Measured  $t$ - $z$  distribution for annihilations obtained with no bias (green circles), left bias (blue triangles), right bias (red triangles) and heated positrons (violet star). The grey dots are from a numerical simulation of antihydrogen atoms released from the trap during the quench. The simulated atoms were initially in the ground state, with a maximum kinetic energy of 0.1 meV. The typical kinetic energy is larger than the depth of the neutral trap, ensuring that all trappable atoms are considered. The 30-ms observation window includes 99% of the 20,000 simulated points. **b**, Experimental  $t$ - $z$  distribution, as above, shown along with results of a numerical simulation of mirror-trapped antiprotons being released from the trap. The colour codes are as above and there are 3,000 points in each of the three simulation plots. In both **a** and **b**, the simulated  $z$  distributions were convolved with the detector spatial resolution, of  $\sim 5$  mm.

In the unlikely event that there are mirror-trapped antiprotons that survive the clearing pulses, it is clear from Fig. 3b that the measured annihilation distributions for the left- and right-bias trapping attempts are not consistent with the model predictions of the drastic deflection and earlier escape of such particles. Nor is the measured no-bias annihilation distribution consistent with the simulation distribution for antiprotons under no-bias conditions. All measured distributions are, however, consistent with the predicted behaviour of neutral antihydrogen (Fig. 3a). In a separate experiment, we intentionally created mirror-trapped antiprotons using extreme potential manipulations, and demonstrated that those that survive the clearing pulses are clearly deflected by the bias fields during the quench, in accordance with the simulations.

The background comprises  $1.4 \pm 1.4$  events (scaled to 335 attempts) detected when trappable antihydrogen is unlikely to be present owing to heating of the positrons, and includes an expected cosmic background of  $0.46 \pm 0.01$  events. As we have shown that the remaining events could not be mirror-trapped antiprotons, we conclude that we have observed the release of antihydrogen atoms that have been magnetically trapped for at least 172 ms.

The extensive diagnostic capabilities (Methods) of the ALPHA device allow us to make an order-of-magnitude theoretical estimate of the expected number of trapped antihydrogen atoms in our experiments. Following the procedure outlined in an earlier work<sup>23</sup>, we estimate that we should detect about 0.4 trapped atoms per attempt, in reasonable agreement with the 0.11 observed here.

We note that although the trapping rate per antihydrogen atom produced is rather low ( $\sim 5 \times 10^{-5}$ , using the overall detection efficiency of about 50%) in our experiment, there is cause for optimism. The parameter space of positron temperature and density—which are the rate-determining factors for our type of mixing—has only begun to be investigated, and the positrons in ALPHA are still warm in comparison with their cryogenic surroundings. The promising technique of evaporative cooling of antiprotons<sup>17</sup> has yet to be used here. Our work is a crucial step towards precision antihydrogen spectroscopy and anti-atomic tests of fundamental symmetries or gravitation.

## METHODS SUMMARY

The ALPHA device has extensive capabilities for characterizing and manipulating charged antimatter plasmas. These include imaging of the plasmas to determine radii and transverse density, temperature measurement by controlled release of the plasma, the rotating-wall technique for control of plasma transverse size and density, evaporative cooling of the positron plasma and autoresonant injection of antiprotons into the positron plasmas.

Extensive simulations of antiproton and antihydrogen motion have been used to inform the experimental programme and to interpret the results of measurements. The simulations track single-particle trajectories using classical force equations.

Event topology is used to distinguish antiproton annihilations from cosmic rays in the silicon detector. The three event characteristics used are the number of reconstructed tracks, the vertex radius, and the deviation from straight-line geometry.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 8 October; accepted 27 October 2010.

Published online 17 November 2010.

1. Dirac, P. A. M. Quantised singularities in the electromagnetic field. *Proc. R. Soc. Lond. A* **133**, 60–72 (1931).
2. Amoretti, M. *et al.* Production and detection of cold antihydrogen atoms. *Nature* **419**, 456–459 (2002).
3. Gabrielse, G. *et al.* Background-free observation of cold antihydrogen with field-ionization analysis of its states. *Phys. Rev. Lett.* **89**, 213401 (2002).
4. Niering, M. *et al.* Measurement of the hydrogen 1S–2S transition frequency by phase coherent comparison with a microwave cesium fountain clock. *Phys. Rev. Lett.* **84**, 5496–5499 (2000).
5. Drobyshev, G. Y. *et al.* Proposal for the AEGIS experiment at the CERN antiproton decelerator (antimatter experiment: gravity, interferometry, spectroscopy). Tech. Report SPSC-P-334; CERN-SPSC-2007-017 (European Organization for Nuclear Research, 2007).
6. Maury, S. The antiproton decelerator: AD. *Hyperfine Interact.* **109**, 43–52 (1997).



7. Bertsche, W. *et al.* A magnetic trap for antihydrogen confinement. *Nucl. Instrum. Methods Phys. Res. A* **566**, 746–756 (2006).
8. Pritchard, D. E. Cooling neutral atoms in a magnetic trap for precision spectroscopy. *Phys. Rev. Lett.* **51**, 1336–1339 (1983).
9. Fajans, J. *et al.* Effects of extreme magnetic quadrupole fields on Penning traps, and the consequences for antihydrogen trapping. *Phys. Rev. Lett.* **95**, 155001 (2005).
10. Andresen, G. *et al.* Antimatter plasmas in a multipole trap for antihydrogen. *Phys. Rev. Lett.* **98**, 023402 (2007).
11. Fujiwara, M. C. *et al.* Particle physics aspects of antihydrogen studies with ALPHA at CERN. *ALP Conf. Proc.* **1078**, 208–220 (2008).
12. Gabrielse, G. *et al.* First capture of antiprotons in a Penning trap: a kiloelectronvolt source. *Phys. Rev. Lett.* **57**, 2504–2507 (1986).
13. Gabrielse, G. *et al.* Cooling and slowing of trapped antiprotons below 100 meV. *Phys. Rev. Lett.* **63**, 1360–1363 (1989).
14. Surko, C. M. & Greaves, R. G. Emerging science and technology of antimatter plasmas and trap-based beams. *Phys. Plasmas* **11**, 2333–2348 (2004).
15. Jørgensen, L. V. *et al.* New source of dense, cryogenic positron plasmas. *Phys. Rev. Lett.* **95**, 025002 (2005).
16. Hess, H. F. Evaporative cooling of magnetically trapped and compressed spin-polarized hydrogen. *Phys. Rev. B* **34**, 3476–3479 (1986).
17. Andresen, G. B. *et al.* Evaporative cooling of antiprotons to cryogenic temperatures. *Phys. Rev. Lett.* **105**, 013003 (2010).
18. Gabrielse, G. *et al.* Antihydrogen production using trapped plasmas. *Phys. Lett. A* **129**, 38–42 (1988).
19. Fajans, J., Gilson, E. & Friedland, L. Autoresonant (nonstationary) excitation of the diocotron mode in non-neutral plasmas. *Phys. Rev. Lett.* **82**, 4444–4447 (1999).
20. Barth, I. *et al.* Autoresonant transition in the presence of noise and self-fields. *Phys. Rev. Lett.* **103**, 155001 (2009).
21. Andresen, G. B. *et al.* Antihydrogen formation dynamics in a multipolar neutral anti-atom trap. *Phys. Lett. B* **685**, 141–145 (2010).
22. Amoretti, M. *et al.* Antihydrogen production temperature dependence. *Phys. Lett. B* **583**, 59–67 (2004).
23. Andresen, G. *et al.* Search for trapped antihydrogen. *Phys. Lett. B*. doi:10.1016/j.physletb.2010.11.004 (in the press).

**Acknowledgements** This work was supported by CNPq, FINEP/RENAFAE (Brazil); ISF (Israel); MEXT (Japan); FNU (Denmark); VR (Sweden); NSERC, NRC/TRIUMF, AIF, FQRNT (Canada); the DOE and the NSF (USA); and EPSRC, the Royal Society and the Leverhulme Trust (UK). We thank them for their generous support. We are grateful to the Antiproton Decelerator team, T. Eriksson, P. Belochitskii, B. Dupuy, L. Bojtar, C. Oliveira, K. Mikluha and G. Tranquille, for the delivery of a high-quality antiproton beam. The contributions of summer students C. C. Bray, C. Ø. Rasmussen, S. Kemp,

K. K. Andersen, D. Wilding, K. Mikkelsen and L. Bryngemark are acknowledged. We would like to thank the following individuals for help: M. Harrison, J. Escallier, A. Marone, M. Anerella, A. Ghosh, B. Parker, G. Ganetis, J. Thornhill, D. Wells, D. Seddon, K. Dahlerup-Pedersen, J. Mourao, T. Fowler, S. Russenschuck, R. De Oliveira, N. Wauquier, J. Hansen, M. Polini, J. M. Geisser, L. Deparis, P. Frichot, J. M. Malzacker, A. Brisswalter, P. Moyret, S. Mathot, G. Favre, J. P. Brachet, P. Mésenge, S. Sgobba, A. Cherif, J. Bremer, J. Casas-Cubillos, N. Vauthier, G. Perinic, O. Pirotte, A. Perin, G. Perinic, B. Vullierme, D. Delkaris, N. Veillet, K. Barth, R. Consentino, S. Guido, L. Stewart, M. Malabaila, A. Mongelluzzo, P. Chiggiato, E. Mahner, A. Froton, C. Lasseur, F. Hahn, E. Søndergaard, F. Mikkelsen, W. Carlisle, A. Charman, J. Keller, P. Amaudruz, D. Bishop, R. Bula, K. Langton, P. Vincent, S. Chan, D. Rowbotham, P. Bennet, B. Evans, J.-P. Martin, P. Kowalski, A. Read, T. Willis, J. Kivell, H. Thomas, W. Lai, L. Wasilenko, C. Kolbeck, H. Malik, P. Genoa, L. Posada and R. Funakoshi.

**Author Contributions** W.B., P.D.B., J.F., M.C.F., J.S.H., N.M. and D.M.S. conceived, designed and constructed the central ALPHA apparatus and participated in all aspects of the experimental and physics programmes. G.B.A., M.D.A., M.B.-R., E.B., S.C., T.F., A.J.H., R.H., M.J.J., A.P., S.S.e.N. and C.S. participated actively in the experimental runs, data taking, on- and offline analysis, and maintenance and modification of the apparatus. D.R.G., A.O. and J.W.S. contributed to all aspects of the detector systems and participated actively in the experimental and analysis efforts. M.C., D.P.v.d.W. and L.V.J. designed and built the positron accumulator and participated in the experimental programme. F.R. performed the particle simulations reported in this paper, made the theoretical estimate of trapping rate and supported the design and experimental programmes with simulations and calculations. P.N. led the design of the ALPHA silicon detector. P.P. was responsible for implementing the silicon detector at CERN and participated in the experimental and analysis programmes. A.D. and A.G. contributed to the experimental shift work. S.J. and J.S.W. contributed theoretical support in the form of atomic or plasma physics calculations and simulations, and provided guidance in the experimental programme. E.S. contributed to the multichannel plate imaging system and participated in the experimental efforts. C.L.C., W.N.H., M.E.H., S.E., S.M. and R.I.T. participated in the experimental programme and the physics planning effort. Y.Y. provided organizational and financial support and participated in physics discussions. L.K. and K.O. provided off-site support for detector electronics and database management systems, respectively. J.S.H. wrote the initial manuscript, which was edited by J.F., M.C.F., P.D.B., N.M. and E.B. before being improved and approved by the entire collaboration.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.S.H. ([hangst@phys.au.dk](mailto:hangst@phys.au.dk)).

## METHODS

**Plasma diagnostics and control.** The rotating-wall technique<sup>24,25</sup> was used several times in each trapping attempt to control the radius and density of both antiproton and positron clouds. The cloud radii and transverse density profiles were measured by releasing the particles onto an imaging detector<sup>26,27</sup> using a multichannel plate coupled to a phosphor screen that was imaged by a charge-coupled-device camera. Equilibrium temperatures were determined by ramping down the axial confining potential and measuring the distribution of escaping particles using either the multichannel plate (positrons) or scintillation detectors (antiprotons). The temperature was obtained from a fit to the high-energy tail of the measured distribution<sup>28</sup>.

We used evaporative cooling<sup>16</sup> to obtain lower positron temperatures. The technique, which we have also recently applied to antiprotons<sup>17</sup>, involves reducing one side of the confining potential well to allow the most energetic positrons to escape. Re-equilibration through collisions results in a lower temperature for the remaining particles. For the trapping experiments described here, the applied, on-axis well depth (neglecting space charge) was reduced from 2.5 to 1.1 V in 500 ms, and about 50% of the initial positrons were lost.

The autoresonant injection of antiprotons into the positron cloud makes use of the fact that the confining potential for the antiprotons is anharmonic, which causes the axial oscillation frequency to decrease with increasing oscillation amplitude. We applied a sinusoidal drive that sweeps downwards through the range of axial frequencies defined by the potential. With a proper choice of drive parameters, the antiprotons autoresonantly lock to the drive frequency and their energies increase as the drive frequency is lowered. Using a drive of  $\sim 55$  mV (on-axis) and a frequency sweep of 350–200 kHz, we were able to inject about 70% of the antiprotons into the positrons in 200  $\mu$ s. This new method of mixing for antihydrogen production was designed to introduce the antiprotons at low longitudinal kinetic energy with respect to the positrons. The initial transverse energy distribution of the antiprotons should also be minimally perturbed by the rapid and precise energy sweep. We note that extensive searches with ATHENA-type mixing<sup>2</sup>, in which the antiprotons were injected into the positrons with several electronvolts of energy, yielded no trapping signal.

**Simulations of antihydrogen and antiproton motion.** We used numerical models to simulate the trajectories of both mirror-trapped antiprotons and trapped antihydrogen atoms as the atom trap was de-energized. The simulations propagate the particles using classical force equations: the Lorentz force for antiprotons and the dipole-gradient force for the antihydrogen atoms. The spatially and temporally varying electric and magnetic fields were included from models of the electrode and magnet geometry. Measurements of the time response of the electrode amplifier chain and calculations of magnetically induced eddy currents were used to reproduce the field dynamics accurately. The simulations model the dynamics after the vast majority of charged particles have been expelled from the trap; thus, the density of particles was low, and single-particle dynamics sufficed. The particles were propagated until they struck the surface of the trap electrodes, whereupon they were considered to have annihilated and we recorded their positions.

**Selection of annihilation events.** Events recorded in the silicon detector can come from cosmic rays and other environmental noise, as well as from the annihilation of antiprotons. Antiproton annihilations on a nucleus produce several charged particles (mostly pions), and they typically produce several tracks in the detector

(Fig. 2a). The radial position,  $r$ , of the reconstructed annihilation vertices was distributed about the inner surface of the electrodes (radius of 22.3 mm). However, our event reconstruction algorithm will typically identify cosmic rays as two back-to-back tracks (Fig. 2b), with the radii of the reconstructed vertices randomly distributed. The environmental noise generally does not register a track or a vertex, and is thus effectively rejected by requiring that each event be associated with a vertex.

To distinguish antiproton annihilation events from cosmic rays and noise background, we used three primary pieces of information about the topology of the events for which our reconstruction algorithm finds a vertex<sup>23</sup>: the number of tracks, the radial position of the reconstructed annihilation vertex and a measure of the deviation of the event topology from that of a straight line passing through the detector. With the third piece of information, compatibility of the event with a cosmic ray is tested by making a linear fit to the hit positions in the event pattern and calculating the sum of the squared residual distances from the fitted line. The antiproton annihilation events tend to give larger values of this ‘squared residual’ than do the cosmic events, which tend to fit well to a straight line.

To optimize the selection criteria, we collected a data sample of cosmic rays ( $\sim 110,000$  events) when there were no antiprotons present in the experiment and we compared this with the sample of antiproton annihilations ( $\sim 170,000$  events) recorded during the mixing phase of the trapping experiments. The mixing phase accumulates data at the maximum read-out rate of the detector ( $\sim 500$  Hz); this rate is large in comparison with the cosmic trigger rate ( $\sim 10$  Hz), so the mixing sample is dominated by annihilations. Following standard practices, we applied ‘cuts’ to the number distributions of the three quantities defined above, to reject cosmic rays while retaining real annihilation vertices. The positions of the cuts were optimized by means of Monte Carlo pseudo-experiments. By performing a large number of pseudo-experiments, we studied the effects of varying the cuts on the resulting significance, averaged over a number of trials. Thus, we derived a set of cuts that would produce, on average, the best statistical significance for cosmic rejection.

The resulting selection criteria for annihilation events were as follows: for two-track events,  $r < 4$  cm and the squared residual was greater than  $2 \text{ cm}^2$ ; for events with three or more tracks,  $r < 4$  cm and the squared residual was greater than  $0.05 \text{ cm}^2$ . With the chosen set of cuts, 99.6% of the cosmic events were rejected, enhancing the signal-to-noise ratio by more than two orders of magnitude while maintaining a high overall efficiency, of 47%, for annihilation detection. To avoid experimental bias, the cuts were optimized using mixing and cosmic data only, and applied a posteriori to trapping search data.

24. Huang, X.-P., Anderegg, F., Hollmann, E. M., Driscoll, C. F. & O’Neil, T. M. Steady-state confinement of nonneutral plasmas by rotating electric fields. *Phys. Rev. Lett.* **78**, 875–878 (1997).
25. Danielson, J. R. & Surko, C. M. Radial compression and torque-balanced steady states of single-component plasmas in Penning-Malmberg traps. *Phys. Plasmas* **13**, 055706 (2006).
26. Andresen, G. B. *et al.* Antiproton, positron, and electron imaging with a microchannel plate/phosphor detector. *Rev. Sci. Instrum.* **80**, 123701 (2009).
27. Andresen, G. B. *et al.* Compression of antiproton clouds for antihydrogen trapping. *Phys. Rev. Lett.* **100**, 203401 (2008).
28. Eggleston, D. L., Driscoll, C. F., Beck, B. R., Hyatt, A. W. & Malmberg, J. H. Parallel energy analyzer for pure electron plasma devices. *Phys. Fluids B* **4**, 3432–3439 (1992).

# Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs

Calvin H. Jan<sup>1,2</sup>, Robin C. Friedman<sup>1,2,3</sup>, J. Graham Ruby<sup>1,2†</sup> & David P. Bartel<sup>1,2</sup>

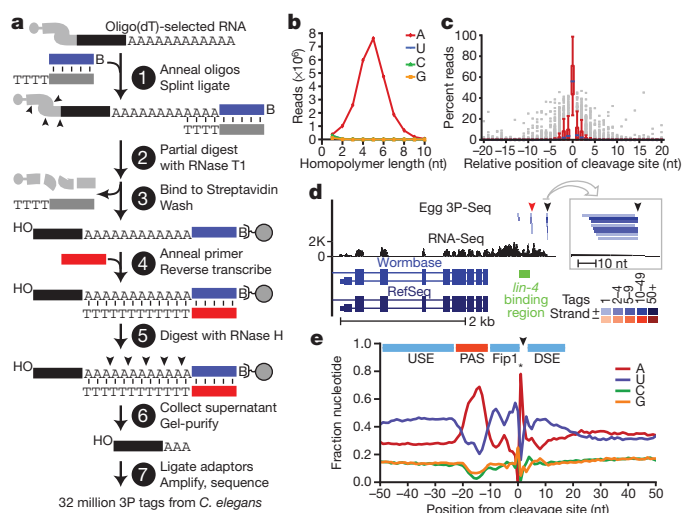
Post-transcriptional gene regulation frequently occurs through elements in mRNA 3' untranslated regions (UTRs)<sup>1,2</sup>. Although crucial roles for 3'UTR-mediated gene regulation have been found in *Caenorhabditis elegans*<sup>3–5</sup>, most *C. elegans* genes have lacked annotated 3'UTRs<sup>6,7</sup>. Here we describe a high-throughput method for reliable identification of polyadenylated RNA termini, and we apply this method, called poly(A)-position profiling by sequencing (3P-Seq), to determine *C. elegans* 3'UTRs. Compared to standard methods also recently applied to *C. elegans* UTRs<sup>8</sup>, 3P-Seq identified 8,580 additional UTRs while excluding thousands of shorter UTR isoforms that do not seem to be authentic. Analysis of this expanded and corrected data set suggested that the high A/U content of *C. elegans* 3'UTRs facilitated genome compaction, because the elements specifying cleavage and polyadenylation, which are A/U rich, can more readily emerge in A/U-rich regions. Indeed, 30% of the protein-coding genes have mRNAs with alternative, partially overlapping end regions that generate another 10,480 cleavage and polyadenylation sites that had gone largely unnoticed and represent potential evolutionary intermediates of progressive UTR shortening. Moreover, a third of the convergently transcribed genes use palindromic arrangements of bidirectional elements to specify UTRs with convergent overlap, which also contributes to genome compaction by eliminating regions between genes. Although nematode 3'UTRs have median length only one-sixth that of mammalian 3'UTRs, they have twice the density of conserved microRNA sites, in part because additional types of seed-complementary sites are preferentially conserved. These findings reveal the influence of cleavage and polyadenylation on the evolution of genome architecture and provide resources for studying post-transcriptional gene regulation.

We developed a high-throughput method to identify 3' ends of mRNAs and other polyadenylated transcripts (Fig. 1a). This method, called poly(A)-position profiling by sequencing (3P-Seq), begins with a splint-ligation that favours ends of poly(A) tails when appending a biotinylated primer-binding site (Fig. 1a, step 1). After partial digestion with T1 nuclease (which cuts after Gs; step 2), the polyadenylated ends are captured (step 3), and the poly(A) tail is reverse transcribed with dTTP as the only deoxynucleoside triphosphate (step 4). Digestion with RNase H releases the polyadenylated ends (step 5), which are purified (step 6) and prepared for high-throughput sequencing (step 7).

3P-Seq was designed to identify the 3' ends of polyadenylated RNAs without recourse to oligo(dT) priming. Oligo(dT) priming can prime on internal A-rich regions of transcripts, thereby yielding artefacts difficult to distinguish from authentic polyadenylated transcripts because the artefacts also have untemplated As<sup>9</sup>. Although untemplated adenylates at the ends of 3P tags could not have arisen from internal-priming artefacts, in principle, such nucleotides could have arisen from polymerase/sequencing errors. Countering this possibility was the observation that homopolymeric runs containing untemplated

nucleotides at the ends of candidate 3P tags were overwhelmingly As (Fig. 1b). Thus, non-genomic terminal adenylates at the ends of 3P tags (a beneficial consequence of incomplete RNase H digestion near duplex termini (Fig. 1a)) provided compelling evidence that they derived from distal ends of bona fide polyadenylated transcripts.

To ensure proper assignment to polyadenylated transcripts, we considered as 3P tags only reads that both mapped uniquely to the genome and possessed at least two 3'-terminal adenylates, of which at least one was untemplated. Nearly 32 million reads from *C. elegans* met these criteria, including millions from each major developmental stage (embryo, L1, L2, L3, L4, adult) as well as dauer L3 worms and germline-deficient *glp-4(bn2)* mutant adults (Supplementary Table 1).



**Figure 1 | Identification of *C. elegans* 3'UTRs.** **a**, Schematic of the 3P-Seq protocol. See text for description. **b**, Sequence composition of homopolymer runs that were found at 3' termini of candidate 3P tags and included  $\geq 1$  untemplated nucleotide. **c**, Cleavage heterogeneity surrounding the most abundant cleavage site (position 0). Box plots show results for 380 cleavage sites that were both between two non-A residues (which enabled precise mapping) and within the top quintile of 3P-tag abundance. **d**, The *lin-14* 3'UTRs. 3P tags from egg were mapped relative to RNA-Seq data<sup>10</sup>, prior mRNA annotations from the indicated databases<sup>6,11</sup>, and the proposed *lin-4*-binding region<sup>4</sup>. Distal and proximal cleavage sites are indicated (black and red arrowheads, respectively). A 50-nucleotide region containing the distal 3P cluster is enlarged (box). Each tag sequence with a unique genome match is depicted as a bar, coloured by tag frequency (key). **e**, Nucleotide sequence composition at mRNA end regions. Shown above are elements implicated in cleavage and polyadenylation (Supplementary Fig. 3c)<sup>30</sup>, with colours reflecting their nucleotide composition (A-rich, red; U-rich, blue). The sharp adenosine peak at position +1 (\*) was due only partly to cleavage before an A. Also contributing to this peak (and to both depletion of A at position -1 and blurring of sequence composition at other positions) was cleavage after an A, for which the templated A was assigned to the poly(A) tail, resulting in a -1 nucleotide offset from the cleavage-site register.

<sup>1</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. †Present address: Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California 94158, USA (J.G.R.).



Microheterogeneity at the cleavage and polyadenylation sites (hereafter called cleavage sites) often produced clusters of related 3P tags (Fig. 1c, d). All tags ending within 10 nucleotides of the most frequently implicated cleavage site were consolidated into a cluster, with this candidate cleavage site representing that of the cluster (Supplementary Data 1). Candidate sites were classified as mRNA cleavage sites if they were bridged by RNA-Seq reads<sup>10</sup> to the stop codons of RefSeq mRNAs<sup>6,11</sup>, as illustrated for *lin-14* (Fig. 1d).

3P-Seq identified 24,033 distinct 3'UTRs, including at least one UTR for 16,261 (83%) of the RefSeq mRNAs (Supplementary Data 2). For 5,331 mRNAs, we revised the longest-isoform annotation by >10 nucleotides (usually by extending it), and for 5,852 mRNAs without 3'UTR annotations, we identified a UTR (Supplementary Fig. 1a). A parallel effort within the modENCODE project used oligo(dT)-based methods to also generate a greatly expanded data set of *C. elegans* 3'UTRs<sup>8</sup>; 8,580 of the 24,033 UTRs identified by 3P-Seq were not identified in that study (Supplementary Table 2). Our data were shared with the modENCODE consortium, thereby enabling them to annotate 8,758 novel UTRs that the oligo(dT)-based methods missed (L. Hillier and R. Waterston, personal communication), Supplementary Table 3). Of the 3,280 RefSeq mRNAs not assigned 3'UTRs using 3P-Seq, most were from predicted genes without evidence of expression (Supplementary Fig. 1b). Of the remainder, most were expressed at extremely low levels (Supplementary Fig. 1c). We estimated that only  $124 \pm 56$  (95% confidence interval) sites were missed by requiring that tags have an untemplated A (Supplementary Fig. 1d). Most histone mRNAs were assigned 3'UTRs, consistent with oligo(dT)-based results<sup>8</sup>, but the polyadenylated forms of these mRNAs did not accumulate to levels detectable on RNA blots (Supplementary Fig. 2).

Apart from the A-rich segment corresponding to the polyadenylation signal (PAS) AAUAAA and its close variants (Supplementary Fig. 3a, b and Supplementary Table 4), the mRNA end regions were U-rich, presumably a feature of the binding sites of factors that enhance cleavage and polyadenylation (Fig. 1e and Supplementary Fig. 3c). Indeed, end regions that lacked a common PAS had exaggerated U-rich features surrounding an A-rich segment located where the PAS normally occurs (Supplementary Fig. 3d), which suggests that appropriate U-rich context can compensate for lack of a strong PAS<sup>12</sup>.

3P-Seq was particularly useful for reliably identifying alternative UTR isoforms. Genes with tandem 3'UTRs possess proximal cleavage sites that, when used, create a shorter UTR that is a subfragment of longer versions (Fig. 2a). When identifying these shorter isoforms, we required that (1) the proximal site be represented by  $\geq 2$  independent 3P tags, (2) that these tags constitute  $\geq 1\%$  of the tags mapping between the distal site and the stop codon, and (3) the site be in an end region non-overlapping with that of a more distal site (that is, that the two cleavage sites be  $\geq 40$  nucleotides apart). These criteria identified 7,795 shorter isoforms, which corresponded to 31% of the Entrez genes with 3P-supported UTRs (Fig. 2a). As expected for sites sometimes

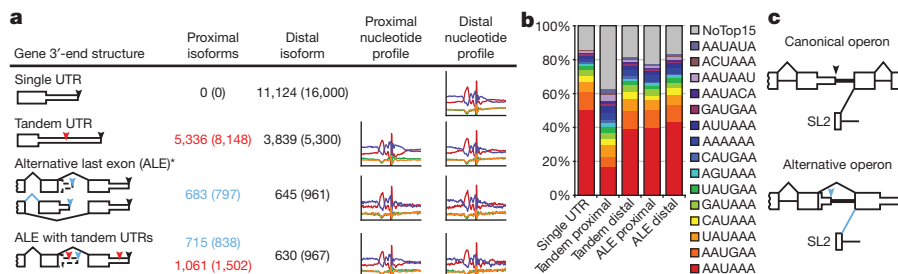
bypassed by the cleavage and polyadenylation machinery to allow production of longer isoforms, a larger fraction lacked a common PAS (Fig. 2b). Although less conserved than PASs for distal-most sites, PASs for proximal sites were more conserved than expected by chance (Supplementary Fig. 4b). Proximal isoforms had lengths typical of *C. elegans* UTRs, whereas distal isoforms were longer than typical UTRs (Supplementary Fig. 4a;  $P < 10^{-300}$ , Wilcoxon rank-sum test), hinting at even more elaborate UTR-mediated regulation.

Oligo(dT)-based results have been interpreted to show that a large class of proximal isoforms lack PASs and instead have A-rich regions immediately following their cleavage sites<sup>8</sup>. 3P-Seq, which avoids oligo(dT) priming, provided no evidence for this novel class of isoforms, suggesting that it is composed of false-positives that arose from internal priming on A-rich UTR regions, as illustrated for the *ubc-18* 3'UTR and confirmed by an RNase-protection experiment (Supplementary Fig. 5a, b). Of the 5,728 proposed cleavage sites supported by oligo(dT)-based methods but not 3P-Seq, 3,900 were sites of putative proximal isoforms (Supplementary Table 3), of which  $\sim 70\%$  seem to have resulted from internal-priming artefacts (Supplementary Fig. 5c).

Genes with alternative last exons (ALEs) generate messages with completely different UTRs (Fig. 2a). We identified 1,398 ALEs distributed across 1,277 Entrez genes. Previous methods identified <25% of these ALEs (Supplementary Fig. 5d), presumably because data acquisition or analyses had focused on regions downstream of annotated stop codons<sup>8</sup>, which illustrates advantages of 3P-Seq for identifying unanticipated UTRs. The PAS motifs and nucleotide composition associated with proximal ALE ends were comparable to those at distal ends (Fig. 2a, b), and the distal isoforms tended to be longer than both proximal isoforms and single UTRs (Supplementary Fig. 4c,  $P < 10^{-5}$  and  $< 10^{-14}$ , respectively, Wilcoxon rank-sum test).

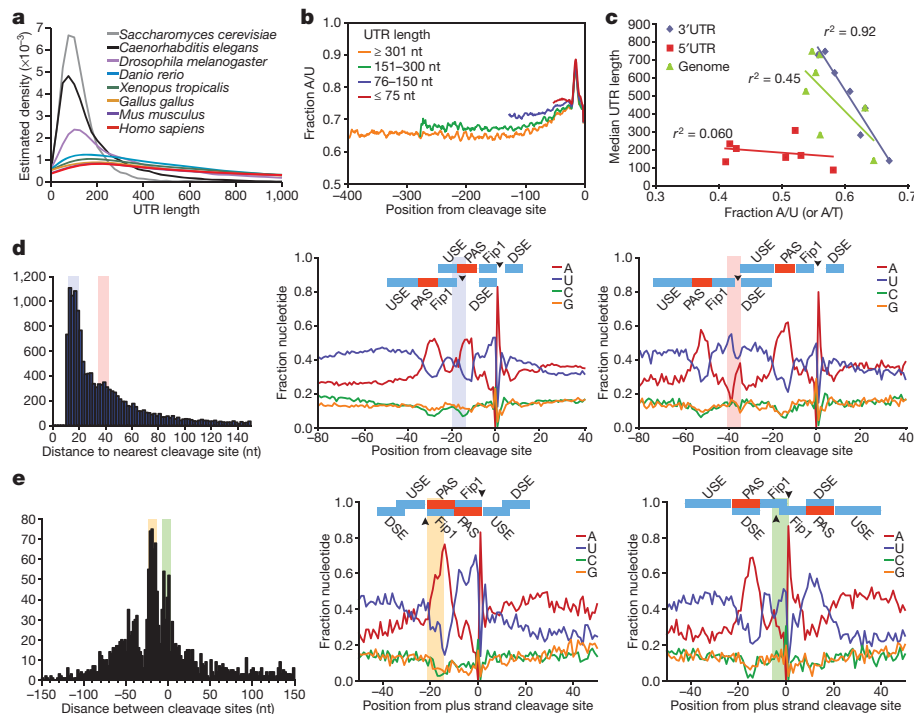
Our analyses also identified a novel gene architecture, called the 'alternative operon'. *C. elegans* operons are each arrays of genes transcribed from a single promoter and split into separate mRNAs through the biochemically coupled processes of 3'-end formation and trans-splicing to splicing leader 2 (SL2)<sup>13</sup>. Reasoning that this coupling could result in SL2 trans-splicing to 3'-splice sites downstream of ALEs, we searched for a gene structure that differed from the canonical operon by a splice junction bridging exons from different genes of an operon (Fig. 2c). This search identified 12 alternative operons, including the *smg-6* locus (Supplementary Fig. 6 and Supplementary Table 5).

Among representative metazoans, *C. elegans* had the shortest 3'UTRs, with a length distribution approaching that of *Saccharomyces cerevisiae* (Fig. 3a) and a median length of 130 nucleotides, only one sixth that of human. *C. elegans* 3'UTRs were also the most A/U rich. Shorter UTRs tended to be the most A/U rich (Fig. 3b), and even after masking the UTR end regions, which are exceptionally U/A rich, a cross-species comparison revealed a significant inverse correlation between 3'UTR length and 3'UTR A/U content ( $P = 0.0003$ ,  $r^2 = 0.92$ , Pearson



**Figure 2 | Alternative 3'UTRs in *C. elegans*.** **a**, Distribution of the 24,033 3P-Seq-supported UTRs among the types of alternative isoforms. For genes with ALEs that have tandem isoforms (bottom), the ALE tally indicates the number of distal isoforms of proximal ALEs (blue) and the tandem tally indicates the proximal tandem isoforms of all ALEs (red). In all cases, the distal isoform is the 3'-most cleavage site for each gene (black arrowhead). Also depicted are proximal tandem sites and proximal ALE sites (red and blue arrowheads,

respectively). Listed (in parenthesis) is the number of cleavage sites associated with each isoform type for the 34,513 3P-Seq-supported cleavage sites (which exceeded the number of unique UTRs because OERs produced multiple cleavage sites for the same UTR). The nucleotide composition near proximal and distal sites is shown (right). **b**, Frequency of PAS motifs for isoform types indicated. **c**, Schematics of canonical and alternative operons.



**Figure 3 | Evolution and topology of 3'-end formation.** **a**, 3'UTR length distributions for the indicated species, considering the most distal annotated isoform for each gene. **b**, A/U content for *C. elegans* 3'UTRs of the indicated lengths. **c**, Relationship between 3'UTR length and 3'UTR A/U content (disregarding content of the last 40 UTR nucleotides), 3'UTR length and genomic A/T content, and 5'UTR length and 5'UTR A/U content for the metazoan species in **(a)** ( $r^2$ , Pearson correlation coefficients). **d**, OERs. Distances between neighbouring cleavage sites are plotted (left). For peaks in the distribution at 15–20 and 35–40 nucleotides (shaded), nucleotide

compositions of OERs are shown (middle and right, respectively), with proposed RNA-recognition elements coloured as in Fig. 1e. Arrowheads indicate cleavage sites, with shading also indicating positions of upstream cleavage. **e**, Convergent UTR overlap. Distances between convergent 3' ends are plotted (left), with negative values indicating overlap. For peaks at 15–22 and (–2)–8 nucleotides of overlap (shaded), nucleotide compositions are shown (middle and right, respectively) as in **(d)**, with shading indicating positions of minus-strand cleavage.

correlation), whereas correlations between either 5'UTR length and A/U content or 3'UTR length and genomic A/T content were less significant ( $P = 0.30$  and  $0.05$ , respectively; Fig. 3c). We speculate that this strong inverse correlation is causal; that is, higher A/U content favours the emergence of A/U-rich motifs that create proximal mRNA ends within existing 3'UTRs, thereby generating progressively shorter UTRs.

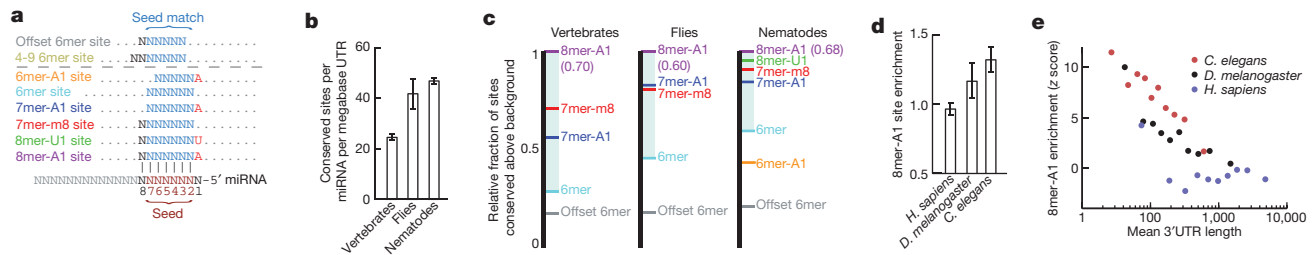
Also potentially related to progressive UTR shortening were the 7,116 UTRs with  $\geq 2$  closely spaced alternative cleavage sites. We did not classify these as tandem UTRs because the cleavage sites were very close to each other ( $< 40$  nucleotides, usually 12–22 nucleotides), implying overlapping end regions (OERs). This overlap tended to be phased, such that U-rich *cis*-acting elements could serve dual functions, binding alternative factors, depending on which cleavage site was being recognized (Fig. 3d). Although previous studies do not distinguish these isoforms from the heterogeneity normally found at UTR ends, proximal and distal OER isoforms were distinct in that each tended to have their own A-rich PASs (Fig. 3d). The 10,480 additional cleavage sites from OERs thus represented the largest class of alternative mRNA isoforms in *C. elegans* (Fig. 2a, compare UTR tallies with cleavage-site tallies).

The few additional nucleotides of distal OER isoforms presumably are dedicated to end recognition and processing (Fig. 3d), leaving little space for regulatory sites that could impart differential regulation. Thus, the importance of the OER isoforms might pertain instead to UTR evolution. The potential of the U-rich regions to serve dual functions would favour the emergence of new cleavage sites with OERs. Moreover, the higher A/U content of *C. elegans* UTRs compared to that of intergenic regions would favour the emergence of more upstream sites than downstream sites, which in turn could lead to progressive UTR shortening as the original signals acquire mutations

rendering them less able to compete for factors (Supplementary Fig. 7a). If nematode UTRs had a propensity to drift towards a minimum UTR length, longer UTRs, which have avoided this shortening, might display more evidence of cleavage-site retention. Indeed, the PASs of long UTRs were more frequently conserved than were those of shorter UTRs (Supplementary Fig. 7b,  $P < 10^{-15}$ , Kolmogorov–Smirnov test).

The alternating U- and A-rich elements defining UTR end regions provided opportunity for motifs to also serve double duty on opposite strands. Indeed, overlap of convergent UTRs occurred with a trimodal distribution peaking at 5, 20 and 40 nucleotides, in which the A-rich PASs of each strand reciprocally served as U-rich motifs of the other strand (Fig. 3e). The bidirectionality of these composite sites was often selectively maintained (Supplementary Fig. 8a). Previously, a single peak in the distribution was observed at  $\sim 20$  nucleotides of overlap, which was attributed to selective pressure to avoid RNAi<sup>8</sup>. Our data indicated a more complex overlap distribution that is better explained by preferential emergence of end regions where end elements of a convergent gene already provide some of the alternating A- and U-rich segments needed for end recognition. Although more extensive overlap can act to enforce mutually exclusive transcriptional regulation<sup>14</sup>, expression of the overlapping gene pairs were no less correlated than were random pairs (Supplementary Fig. 8b). Hence, gene topology using palindromic arrangement of bidirectional elements provides a mechanism for genome compaction, effectively minimizing intergenic space downstream of 2,448 genes (a sixth of all genes with 3P-Seq-identified ends) without significantly impacting their regulatory autonomy.

Before considering targeting of the newly annotated 3'UTRs by microRNAs (miRNAs), we updated the set of confidently identified miRNAs using  $\sim 23$  million genome-matching small-RNA sequences<sup>15</sup>.



**Figure 4 | MicroRNA targeting.** **a**, Expanded repertoire of seed-matched sites preferentially conserved in nematode 3'UTRs. Sites conserved only marginally above chance are above the dashed line. Watson–Crick-matched residues, blue or black; residues independent of the miRNA sequence, red. **b**, Density of miRNA sites conserved above background, combining all site types at the maximally sensitive cutoff. Error bars, one standard deviation (calculated by repeating the analysis for each site type 50 times, each time using a different cohort of control sequences that matched the properties of the miRNA sequences<sup>18</sup>). **c**, Relative strength of miRNA site types across clades. Within

Methods shown to identify miRNAs reliably in mammals<sup>16</sup> provided confident support for 147 annotated genes and 12 additional genes (Supplementary Table 6, Supplementary Fig. 9 and Supplementary Text). Five of the newly identified miRNAs derived from mirtrons (Supplementary Table 6), which are spliced and debranched introns that fold into pre-miRNA hairpins, thereby bypassing Drosha processing<sup>17</sup>. Although mirtrons are typically thought to be spliced from pre-mRNAs, three newly identified mirtrons and two pre-miRNAs reclassified as mirtrons (*mir-255* and *mir-2220*) are derived from host transcripts that did not seem to be protein coding (Supplementary Fig. 10 and Supplementary Table 6). We also generated developmental expression profiles for the 159 confidently annotated genes (Supplementary Fig. 11, Supplementary Tables 6 and 7 and Supplementary Text).

Methods used previously to detect miRNA site conservation in vertebrate genome alignments<sup>18</sup> found six types of preferentially conserved sites that matched the miRNA seed region (Fig. 4a), including an octamer site (8mer-U1) and a hexamer site (6mer-A1) not observed in vertebrates (Supplementary Figs 12a–c, 13 and Supplementary text). Efficacy of these six types was confirmed using two large-scale experimental data sets: mRNA fragments crosslinked to *C. elegans* miRNA silencing complexes<sup>19</sup> and mRNA changes in *C. elegans* miR-124 mutants<sup>20</sup> (Supplementary Fig. 12d and Supplementary Table 8).

Summing results for all six site types indicated that *C. elegans* UTRs have at least  $9,093 \pm 146$  (95% confidence interval) selectively maintained miRNA sites, and that at least  $27.4\% \pm 4.8\%$  of the *C. elegans* 3'UTRs have been under selective pressure to retain miRNA targeting (Supplementary Fig. 12b, c). This percentage was nearly threefold greater than that detected previously in nematodes<sup>21</sup> and about half that observed for human UTRs<sup>18</sup>, despite substantially shorter lengths of nematode UTRs, fewer nematode genomes available, and fewer conserved miRNA families in nematodes (60, compared to 87 in vertebrates). As in vertebrates<sup>18</sup>, few preferentially conserved sites had mismatches or wobbles to the seed nucleotides (Supplementary Fig. 14). Indeed, the three most compelling sites with seed mismatches (two *let-7* sites in *lin-41* and one *let-7* site in *hbl-1*; Supplementary Table 9) had all been implicated by earlier genetic studies<sup>22,23</sup>. The updated miRNA target predictions will be presented in TargetScanWorm, release 5.2 (targetscan.org).

Compared to human 3'UTRs, *C. elegans* 3'UTRs had twice the density of selectively conserved miRNA sites (Fig. 4b and Supplementary Fig. 15d). This difference was attributed partly to the two additional site types conserved in nematodes and partly to the higher fractions of hexamer and heptamer sites preferentially conserved in nematodes (Fig. 4c). *Drosophila*, which has intermediate 3'UTR lengths (median 224 nucleotides), had intermediate fractions of sites

each clade, two species of comparable divergence were selected. For each miRNA site type, the fraction of sites conserved above background in the two species was normalized to that of the 8mer-A1 (shown in parentheses). **d**, Enrichment of 8mer-A1 3'UTR sites above expectation based on dinucleotide content. Error bars, one standard deviation, derived as in (b). **e**, Relationship between 3'UTR length and site enrichment. Site enrichment is plotted for 3'UTRs of the indicated species sorted by length into ten equally sized bins.

conserved (Fig. 4c). Because the relative conservation of site types correlates well with their efficacy<sup>18</sup>, species with shorter 3'UTRs presumably have increased relative efficacy of site types that impart marginal repression in vertebrates, such as most hexamer sites. With this increased miRNA targeting promiscuity, *C. elegans* could cope with shorter UTRs without sacrificing as much miRNA-mediated regulation.

MicroRNA sites were enriched in *C. elegans* 3'UTRs irrespective of conservation ( $P < 10^{-3}$ , binomial test; Fig. 4d). This enrichment was not observed in other regions of *C. elegans* mRNAs nor for human miRNA sites in any region of human mRNAs (Supplementary Fig. 15a). Perhaps in humans, the evolutionary depletion of detrimental miRNA sites balances the selective retention of beneficial sites, whereas in *C. elegans*, with its short UTRs, the depletion is not sufficient to balance the selective retention of beneficial sites (Supplementary Text). In this model, miRNA site enrichment would be a property of short 3'UTRs in any context. Indeed, enrichment of miRNA sites inversely correlated with mean 3'UTR length in both interspecies and intraspecies comparisons (Fig. 4d, e). Increased miRNA site density and increased efficacy of marginal site types in the context of short 3'UTRs are both likely to generalize to other *cis*-regulatory elements. Indeed, in *C. elegans*, the ~6,400 tandem UTR events occurred at one event per 560 nucleotides, a density over five times that reported in human UTRs<sup>24,25</sup>.

3P-Seq provided a more comprehensive and reliable view of *C. elegans* 3'UTRs and the basis for insights into their formation, evolution, and regulation. The method should provide analogous results when applied to other eukaryotes with poorly annotated 3'UTRs, that is, most sequenced eukaryotes. 3P-Seq should also be informative for human studies, where it could shed light on shorter UTR isoforms, including those associated with cell proliferation and oncogenic transformation<sup>26–28</sup>.

## METHODS SUMMARY

Nematodes were grown and RNA isolated as described<sup>29</sup>. 3P-Seq was performed as outlined in Fig. 1a. Reads that both mapped to a single locus in the genome and possessed  $\geq 2$  3'-terminal adenylates ( $\geq 1$  untemplated) were carried forward as 3P tags. Tags were iteratively clustered into representative cleavage sites and bridged to transcript models with RNA-Seq data (accession SRA003622.7)<sup>10</sup>. Poly(A) signals were identified as hexamers with position-dependent enrichment similar to AAUAAA. Cleavage sites 5' of terminal exons indicated ALEs. For each last exon, cleavage sites mapping between the stop codon and the 3'-most 3P-Seq-supported cleavage site indicated tandem isoforms. Conservation analysis was as described<sup>18</sup>, except five UTR conservation bins were used for *D. melanogaster* and four were used for *C. elegans*, in order to compensate for the smaller total sequence space of 3'UTRs in these species. For comparisons between mammals, flies and nematodes, pairs of species were chosen such that each had a 3'UTR nucleotide-level divergence rate of ~0.55.



**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 20 September; accepted 29 October 2010.**

**Published online 17 November 2010.**

1. Moore, M. J. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**, 1514–1518 (2005).
2. Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
3. Ahringer, J. & Kimble, J. Control of the sperm-oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3' untranslated region. *Nature* **349**, 346–348 (1991).
4. Wightman, B., Burglin, T. R., Gatto, J., Arasu, P. & Ruvkun, G. Negative regulatory sequences in the *lin-14* 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. *Genes Dev.* **5**, 1813–1824 (1991).
5. Merritt, C., Rasoloson, D., Ko, D. & Seydoux, G. 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Curr. Biol.* **18**, 1476–1482 (2008).
6. Rogers, A. *et al.* WormBase 2007. *Nucleic Acids Res.* **36**, D612–D617 (2008).
7. Mangone, M., Macmenamin, P., Zegar, C., Piano, F. & Gunsalus, K. C. UTRome.org: a platform for 3' UTR biology in *C. elegans*. *Nucleic Acids Res.* **36**, D57–D62 (2008).
8. Mangone, M. *et al.* The landscape of *C. elegans* 3' UTRs. *Science* **329**, 432–435 (2010).
9. Nam, D. K. *et al.* Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc. Natl Acad. Sci. USA* **99**, 6152–6156 (2002).
10. Hillier, L. W. *et al.* Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res.* **19**, 657–666 (2009).
11. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65 (2007).
12. Nunes, N. M., Li, W., Tian, B. & Furger, A. A functional human poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J.* **29**, 1523–1536 (2010).
13. Evans, D. *et al.* A complex containing CstF-64 and the SL2 snRNP connects mRNA 3' end formation and trans-splicing in *C. elegans* operons. *Genes Dev.* **15**, 2562–2571 (2001).
14. Prescott, E. M. & Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. *Proc. Natl Acad. Sci. USA* **99**, 8796–8801 (2002).
15. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* **31**, 67–78 (2008).
16. Chiang, H. R. *et al.* Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* **24**, 992–1009 (2010).
17. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83–86 (2007).
18. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
19. Zisoulis, D. G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nature Struct. Mol. Biol.* **17**, 173–179 (2010).
20. Clark, A. M. *et al.* The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans*. *Nucleic Acids Res.* **38**, 3780–3793 (2010).
21. Lall, S. *et al.* A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* **16**, 460–471 (2006).
22. Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
23. Abrahante, J. E. *et al.* The *Caenorhabditis elegans* *hunchback*-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. *Dev. Cell* **4**, 625–637 (2003).
24. Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33**, 201–212 (2005).
25. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
26. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
27. Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. USA* **106**, 7028–7033 (2009).
28. Mayr, C. & Bartel, D. P. Widespread shortening of 3' UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673–684 (2009).
29. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
30. Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3'-end processing. *Cell Mol. Life Sci.* **65**, 1099–1122 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank C. Burge and members of the Bartel lab for discussions and the WIBR Genome Technology Core for sequencing. This work was supported by NIH grant GM067031 (D.P.B.), a National Science Foundation predoctoral fellowship (C.H.J.) and a Krell Institute/Department of Energy Computational Sciences Graduate Fellowship (R.C.F.).

**Author Contributions** C.H.J. performed the experiments and computational analyses of 3P-Seq data. R.C.F. performed the computational analyses of miRNA targeting and motif conservation. J.G.R. performed the computational analyses of miRNAs. All authors contributed to study design and manuscript preparation.

**Author Information** 3P-Seq reads and 3P tags were deposited at the GEO as fastq and BED files, respectively (GSE24924). MicroRNA genes were deposited at miRBase (miR4805–miR4816). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.P.B. ([dbartel@wi.mit.edu](mailto:dbartel@wi.mit.edu)).

## METHODS

**3P-Seq libraries.** Nematodes were grown and RNA isolated as described<sup>29</sup>. For each library, 30 µg total RNA was enriched for polyadenylated mRNA (Dynabeads Oligo(dT)<sub>25</sub>, Invitrogen). Enriched RNAs were ligated to a 5'-phosphorylated, 3'-biotinylated oligonucleotide adaptor (p-agcguguagggcaccacGACATAC-Biotin; lowercase, RNA; uppercase, DNA) using the splint DNA oligonucleotide (ATGGTGGCCCTACACGCTTTTTTTT) and 1 U Rnl II RNA ligase (New England Biolabs) in a 20 µl reaction for 16 h, according to the manufacturer's instructions. After phenol extraction and precipitation, the RNA was partially digested with 3 U RNase T1 (biochemistry grade, Ambion) in a 100 µl reaction for 20 min at 22 °C, and 3' fragments were captured with 100 µl streptavidin-coated beads (Dynabeads M-280 streptavidin, Invitrogen) in 400 µl B buffer (5 mM Tris-Cl, pH 7.5, 0.5 mM EDTA, 1 M NaCl) for 15 min rotating at room temperature (~22 °C). After one wash in B buffer, beads were washed twice in 400 µl W buffer (10 mM Tris-Cl, pH 7.5, 1 mM EDTA, 50 mM NaCl) at 50 °C, then equilibrated in reverse transcription buffer (Invitrogen). The reverse transcription primer (GTATGTGCATGGTGGCCCTACACGCT) was annealed and then extended with dTTP as the only deoxynucleoside triphosphate, using 1 U reverse transcriptase (Superscript III, Invitrogen) in 25 µl for 20 min at 48 °C. Polyadenylated RNA fragments were released into solution by adding 1 U RNase H (Invitrogen) and digesting for 25 min at 37 °C. After precipitation, fragments were ligated to a pre-adenylated 3' adaptor (AppAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-C3spacer, synthesized as in ref. 29) with 10 U T4 RNLI (NEB) in a 10 µl reaction for 2 h at 22 °C, and ligation products were gel-purified (excising 75–300 nucleotide RNAs) and prepared for Illumina sequencing with a protocol used for strand-specific mRNA-Seq<sup>31</sup>. Because sequencing started at the residues corresponding to the 3' ends of the RNA fragments, which for the 3P tags were all adenylates, cluster definition was somewhat compromised, which lowered the yield of 3P tags. In experiments performed after those described here, we obtained higher yields of 3P tags when defining the clusters from images in the middle of the run by starting the Illumina base calling with the middle images and then reading the images from the first part of the run after the clusters had been defined. A detailed 3P-Seq protocol is available at [web.wi.mit.edu/bartel/pub/protocols.html](http://web.wi.mit.edu/bartel/pub/protocols.html).

**Distal mRNA cleavage sites.** The reverse complements of the sequencing reads were considered candidate 3P tags. These candidate tags were aligned to the *C. elegans* genome (WS190) with Bowtie<sup>32</sup>, using alignment parameters '-q --solexaquals -5 3 -1 25 -n 1 -e 240 -m 1' to allow for the presence of untemplated nucleotides at their 3' termini. Sequences that both mapped to a single genomic locus and possessed  $\geq 2$  3'-terminal adenylates, of which at least one was untemplated, were carried forward as 3P-Seq tags. The most 3'-terminal non-adenosine base of each tag was considered a candidate cleavage site. Genomic loci were then marked off, using a set of RefSeq transcripts with non-redundant 3' UTRs, with each locus corresponding to the region between the annotated 5' terminus of the transcript and the annotated 5' terminus of the downstream gene on the same genomic strand. Candidate sites mapping to each locus were sorted from most abundant to least, with equally abundant sites ordered 3'-most first. Clusters were then built from all sites within a 21-nucleotide window centred on the site with the most tags (combining data from all libraries). This process was iterated until all 3P tags were assigned to clusters (with some clusters containing only one tag). The central site of each cluster was then evaluated as a potential mRNA cleavage site using RNA-Seq data (accession number SRA003622.7)<sup>10</sup>. The number of RNA-Seq reads covering each base of the transcript was tallied, and a 50-nucleotide window was slid from the stop codon to the candidate terminus, after masking bases contained within annotated introns. A candidate site was assigned to an upstream protein-coding region if the median per-base RNA-Seq coverage in all windows was above 0. These sites were filtered further, requiring that the median per-base RNA-Seq coverage in the implied UTR was  $\geq 5\%$  that of the corresponding protein-coding region and that the maximum per-base RNA-Seq coverage in the UTR did not exceed five times that of the coding-region maximum. Among the sites that passed these filters, the distal cleavage site of a gene within the locus was the site of the 3'-most cluster that contributed  $\geq 1\%$  of the 3P tags from the locus.

**Poly(A) signals.** Genes with single UTRs were used to search for position-dependent enrichment of hexamer motifs near the cleavage site. To establish the region where PASs were expected to occur, AAUAAA enrichment was analysed at each position within the 50 nucleotides upstream of cleavage sites. At each position, significance was determined by the binomial test against the first-order Markov expectation for AAUAAA (Supplementary Fig. 3). The region with significant AAUAAA enrichment (9–25 nucleotides upstream of the cleavage site) was analysed for enrichment of other hexamers, after removing the UTRs with AAUAAA in the region. The most significantly enriched alternative hexamer was identified as above, and sequences containing this hexamer were removed and the process was iterated another 13 times. Enrichment analysis was also performed by an alternative process, in which the

first-order Markov expectations were replaced with the hexamer frequencies in an equally wide control window starting 50 nucleotides upstream of the cleavage site, and significant enrichment was determined by the Fisher's exact test. PASs were assigned to each cleavage site by searching the region 9–25 nucleotides upstream of the site, considering the 15 most significantly enriched hexamers (Supplementary Fig. 3 and Supplementary Table 4) and in cases of matches to more than one hexamer, giving preference to the one most significantly enriched in the global analysis that calculated enrichment using upstream control sequences.

**Proximal alternative sites.** For each Entrez gene, candidate cleavage sites (as defined above) were considered as proximal alternative cleavage sites if they (1) mapped between the 5'-most end and the 3'-most 3P-supported cleavage site of the gene, (2) were from clusters containing  $\geq 1\%$  of the tags from the gene, and (3) were from clusters containing two independent 3P-tags. Tags were considered independent if they either (1) were sequenced in independent libraries, (2) mapped to different cleavage sites, or (3) mapped to the same cleavage site but had different numbers of terminal adenylates. For each RefSeq transcript, proximal alternative sites 3' of the stop codon were classified as proximal tandem sites. Candidate ALEs were identified by proximal alternative sites that mapped internally to genes, excluding the exons with any nucleotides 3' of the distal-most stop codon for each Entrez gene annotation<sup>33</sup>. Identification of ALEs was particularly challenging in *C. elegans* for two reasons. First, many gene annotations had limited experimental validation. Second, *C. elegans* has a high density of genes<sup>34</sup>, at least 15% of which are organized as operons<sup>35</sup>. Identification of ALEs thereby depended on experimental validation of the exons as alternative, that is, validation that the cleavage site is sometimes removed due to alternative splicing. ALEs were required to have the support of one of the following: (1) an EST omitting the cleavage site and aligning to both an upstream and downstream exon relative to the cleavage site; (2) an RNA-Seq read spanning the exon junction between the upstream and downstream exons; (3) 3P tags mapping to exons downstream of the ALE. In addition, ALEs were required to have an in-frame stop codon before the cleavage site. If no appropriate stop codon was annotated for novel ALEs, the nearest upstream exon was extended to the ALE cleavage site and the first in-frame stop codon was used.

**Experimental evaluation of cleavage sites.** Probes for RNase-protection experiments were designed to span proximal and distal cleavage sites of genes identified as having tandem UTRs either by both 3P-Seq and oligo(dT)-based methods (*rpl-12*, *kin-19*) or by only oligo(dT)-based methods (*ubc-18*)<sup>8</sup>. Templates for T7 transcription were amplified from N2 bristol genomic DNA using the following primer pairs: GAACAGCCCCAATCCGTTGG, CAACACCAGTGTCTTTTCGATAC (*rpl-12*); CTCCTTTTGGCTCCAAATGCC, AGGGTGTACGGGAAATAGC (*kin-19*); GGAGCACACTCGAAAGCACG, CCGTGTGTGTATCGGCAACATC (*ubc-18*). Amplicons were cloned into a vector suitable for T7 transcription. Probes were body-labelled during *in vitro* transcription (MAXIscript, Ambion) and gel-purified on denaturing 5% acrylamide gels. RPAs were performed with 10<sup>5</sup> counts per minute probe and 15 µg total RNA, hybridized overnight at 42 °C and digested for 45 min in a 1:50 dilution of RNase A/T1 at 22 °C (RPA III, Ambion). Products were resolved on denaturing 5% acrylamide gels and visualized by phosphorimaging.

**Conservation analyses.** 3' UTR alignments were extracted from Multi-Z alignments, (6-way for nematodes, 15-way for *Drosophila*) from the UCSC genome browser<sup>36</sup>, starting with *D. melanogaster* RefSeq annotations for *Drosophila* UTRs. Conservation analyses were as described<sup>18</sup>, except that five UTR conservation bins were used for *D. melanogaster* and four were used for *C. elegans*, to compensate for the smaller total sequence space of 3' UTRs in these species. Analyses of miRNA sites considered 60 *C. elegans* miRNA families with nucleotides 2–8 conserved throughout the *Caenorhabditis* clade (Supplementary Table 7) and 51 *Drosophila* miRNA families with nucleotides 2–8 conserved to *Drosophila pseudoobscura* (Supplementary Table 10). For imperfect site types, only the position of the bulged or mismatched nucleotide needed to be conserved, not the nucleotide itself. For comparisons between mammals, flies and nematodes (Fig. 4; Supplementary Fig. 15) pairs of species analysed were *Homo sapiens* and *Monodelphis domestica*, *Drosophila melanogaster* and *Drosophila willistoni*, and *Caenorhabditis elegans* and *Caenorhabditis briggsae*, each of which had a 3' UTR nucleotide-level divergence rate of ~0.55.

**k-mer enrichment.** For each type of miRNA site, 1,000 cohorts of control *k*-mers were chosen to match the site length, number of G + C nucleotides, and number of CpG dinucleotides. Enrichment was calculated by comparing the number of site occurrences in the region of interest to the mean number of occurrences for the controls. For each *k*-mer, the expected occurrence was determined using a first-order Markov model, and the *P* value was the fraction of control cohorts with ratios of observed-to-expected occurrences more extreme than that of the sites. When analysing enrichment in *Drosophila* and human 3' UTRs (Fig. 4 and Supplementary Fig. 15), only RefSeq annotations with 'validated' status were used.

31. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).

32. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
33. Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* **35**, D26–D31 (2007).
34. The C elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
35. Blumenthal, T. Trans-splicing and operons. *WormBook* **25**, 1–9 (2005).
36. Karolchik, D. et al. The UCSC genome browser database. *Nucleic Acids Res.* **31**, 51–54 (2003).